

Think global

'Virtual globe' software is transforming our ability to visualize and hypothesize in three dimensions. Educators take note.

Millions of people across the world are zooming in from space, flying across continents, and swooping over mountains and through cities, thanks to Google Earth, NASA's World Wind and other free virtual globes.

The ability to model the Earth in exquisite three-dimensional detail was previously only approached on the desktops of professional users of geographical information systems (GIS). But even they were unable to publish high-resolution globes on the Internet, because of the sheer volume of the data — a globe with a resolution of one metre would take years to download using even a fast Internet connection. Virtual globes overcome this problem with elegant engineering, using a tiling structure that sends progressively higher-resolution data as one zooms in. This and other tricks drastically reduce the size of file transfers, and allow visualization with almost zero latency on a decent broadband connection.

Scientists are already experimenting with these tools to showcase their research to the public in visually appealing ways and to speed responses to natural disasters (see pages 776 and 787). Ultimately, such accurate digital representations promise to anchor and unify much digital information about the Earth, while also helping to integrate the efforts of researchers from many disciplines.

Rita Colwell, a microbiologist and former head of the National Science Foundation, has described GIS as the "ultimate, original, multidisciplinary language". Her own research is a shining example. Realizing that cholera epidemics spread inland from the coast, she correlated them with seasonal plankton blooms, discovering on the way that the *Vibrio cholerae* bacteria that cause cholera associate with gravid copepods, helping to break open their egg sacs by secreting chitinases. She went on to use remote sensing for a global predictive system for epidemics. As she has said, a major need is "to appreciate the complex reactions that characterize ecosystems — it is too complex for any one discipline".

By making it child's play to share and view multiple large data sets, virtual globes lower the barrier to entry for scientists with little GIS

experience. Visualization itself can lead to new insights. Having tasted such visualizations, many researchers will be tempted to go beyond them to exploit the full capacities of GIS science to analyse vast arrays of disparate data in their spatial context.

The opportunities and power of GIS are expanding rapidly because of converging technological trends. The quality of spatial and remote-sensing data is sharply increasing in many fields, as are data-mining techniques, which can help lead to new hypotheses. Mobile global positioning system (GPS) devices are raising the prospect of collecting location-specific information quickly and cheaply, making it possible for large networks of intelligent devices to map and monitor a release of toxic gas, for example, and predict its spread.

To meet such spatial opportunities, researchers and students will need training in spatial sciences. The risks of using computational packages as 'black boxes' are well known, and are even greater, if anything, with GIS. As GIS experts have noted: "The production of visually appealing, even statistically sound, results that do not reveal anything useful about either pattern or process is perhaps the greatest danger facing newcomers to this powerful technology" (*Nature Rev. Microbiol.* **1**, 231–237; 2003). It is therefore encouraging to note that last week the US National Academy of Sciences called for the introduction of "spatial thinking", including GIS, into school curricula, and for government research agencies to launch research into the nature of the cognitive processes involved in such thinking.

Many outstanding minds, including Einstein, Faraday, Kekulé and Heisenberg, have attributed their key insights to the ability to think spatially. Let's hope that the upcoming generation of three-dimensional gamers and Google Earthers will yield even more spatially adept prodigies ready to confront global challenges. ■

"Scientists are already experimenting with these tools to showcase their research to the public and to speed responses to natural disasters."

It's academic

The development of scientific academies could help to put science to work in Africa.

The medical, scientific and environmental challenges facing the continent of Africa can seem simply overwhelming. Some of them, such as the provision of health care and basic scientific education, are bound to be expensive to address. But others could be tackled by less expensive and more subtle means: the development of properly functioning scientific academies in African nations.

The US National Academy of Sciences, under its former president Bruce Alberts, began a laudable, long-term project to build up the prowess of such academies in three African nations: Nigeria, South Africa and Uganda. But as a meeting in Amsterdam of African academy officials heard earlier this month, rapid progress could be made in other countries too, if scientists, government officials and donor nations would step up to the plate.

Self-elected bodies of accomplished scientists make a significant contribution to public discourse in most wealthy nations. In Britain, the Netherlands and the United States, for example, academies have long played a broadly productive role in raising the level of public discussion on technical issues. They also provide governments



with advice on matters as diverse as public transport and nuclear-waste disposal.

There is an opportunity, in the rapidly developing polities of modern Africa, for similar academies to emerge there. The existing academies have had little influence. They were not involved, for example, in shaping the New Partnership for Africa's Development (NEPAD), an important collaboration between Africa's governments whose plans include the creation of research centres and networks to tackle continent-wide woes, such as malaria and poor water supply.

This lack of clout has been equally apparent at the national level. In 2002, for example, when Zambia decided to reject food aid from the United States on the grounds that it contained genetically modified maize, its president Levy Mwanawasa appealed on television for the country's scientists to help him out. The broadcast was the first Zambia's science academy knew of his interest, as there was no established channel for communication between the academy and his government in Lusaka.

There is no good reason why this state of affairs should persist. Academies are not particularly expensive to run, and they already exist in many African countries. Although their membership is small and their expertise is in many cases limited, there are areas in which they could be helping their respective governments right now. What is often missing are the political skills needed to get such advice heard, in government or in the media.

The participants in the National Academy's ten-year project, supported by \$20 million from the Bill & Melinda Gates Foundation, are

already showing the way ahead, particularly with regard to health issues. In Uganda, for example, scientists will be trained to work effectively with media outlets to get their message across, and regular meetings will be arranged between researchers, businesses, government officials and politicians to discuss their responses to malaria.

For this model to make headway, of course, political leaders need to be convinced of its value. That will only happen in nations with leaders who are genuinely accountable to their electorates, and where a reasonably free press permits public discussion of the issues.

Progress on both these fronts remains patchy, but scientists can strengthen their prospects of being heard by working to build up the effectiveness of their own academies, while finding proactive means of getting their message out to politicians, the media and the wider public. Greater expertise will then be brought to bear, particularly with regard to pressing African issues, such as standards in education at all levels.

Even in the best of circumstances, it is extremely difficult to get politicians to listen to independent scientific advice. African capitals do not enjoy the best of circumstances. Yet the political situation in many of these capitals is fluid, and determination from scientists to make their collective voices heard could go a long way towards bringing their expertise to bear on African problems. ■

"Determination from scientists to make their voices heard could go a long way towards bringing their expertise to bear on African problems."

NASA in reverse

The US space agency's relationship with scientists is hitting a new low.

The NASA public-affairs office has this month been accused of trying to censor one of its most eminent climate researchers, James Hansen of the Goddard Institute for Space Studies in New York. And more evidence of press releases being doctored for political ends at the space agency is likely to emerge, disturbing everyone who values the free flow of scientific information to the public.

But the Hansen debacle is just one element of the increasingly adversarial relationship that is developing between NASA and the research community. The sour mood was apparent at last month's American Astronomical Society meeting in Washington DC, when NASA's science chief Mary Cleave told assembled scientists that her most important "stakeholders" were the White House and Congress. Cleave's real (if unintentional) message was clear: don't expect NASA to advocate research, as we work for other interests.

Scientists were also dismayed at how fast NASA administrator Mike Griffin reneged on a promise made last autumn not to take "one thin dime" from space science to address the budget problems of the space shuttle and the space station. At his budget news conference on 6 February, Griffin confessed to doing just that, shifting \$2 billion over five years from research to the astronaut programme.

The cuts to science were deep, and they were decided behind closed doors. Take the research and analysis grants that fund the

basic intellectual work underlying NASA's space missions. Previous NASA administrators, recognizing that many space scientists rely on these grants to stay in business, kept the grant programme healthy. But the new budget slashes research grants by 15–25%, and by even more in areas such as astrobiology. And NASA is yet to give details of how deep the cuts actually are.

Nor did the agency bother to inform some scientists that their projects were being axed. One investigator who was awaiting final approval of her NuSTAR Explorer mission after two years of hard work heard Cleave announce its cancellation at a press conference. Such a cavalier approach to communication is prompting outrage in the community.

NASA is undergoing a historic shift in direction without consulting scientists or paying attention to their advice. Projects with great appeal to scientists and to the public — including the search for planets around other stars and the study of dark energy — are being abandoned so that NASA can return astronauts to the moon half a century after the Apollo landings.

Griffin still enjoys some good will from researchers who know him. They understand that his overall budget is set by the White House, and that he is only cutting science reluctantly. But that good will is soon going to vanish if the Bush administration continues to steer the nation's \$17-billion space programme on such a controversial course, without listening to alternative views of what NASA should be doing. ■

"NASA is undergoing a historic shift in direction without consulting scientists or paying attention to their advice."

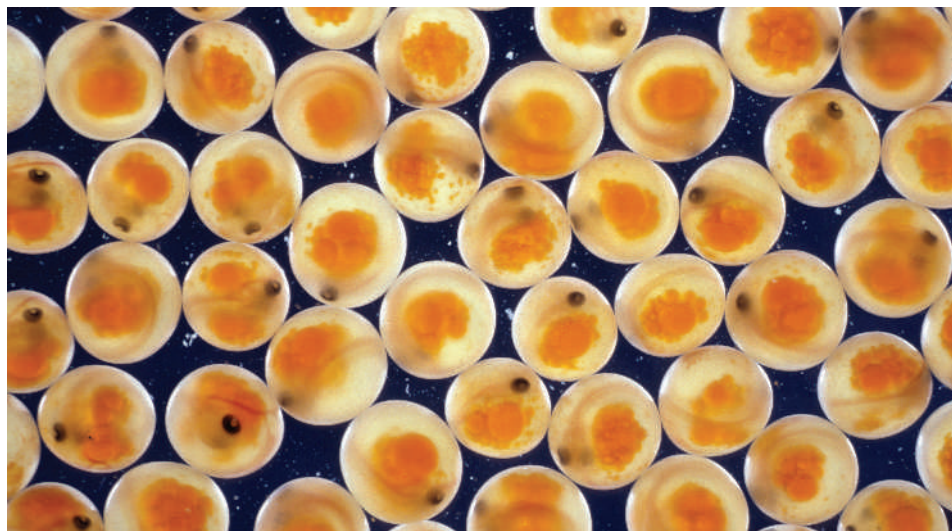
RESEARCH HIGHLIGHTS

Male cells make eggs*Proc. Natl Acad. Sci. USA*

doi:10.1073/pnas.0509218103 (2006)

Cells that normally manufacture sperm can churn out working eggs instead, say researchers — at least in rainbow trout (*Oncorhynchus mykiss*).

One population of stem cells in the embryo ultimately gives rise to either eggs or sperm. Goro Yoshizaki of the Tokyo University of Marine Science and Technology, Japan, and his colleagues tested whether cells in the adults' testes retain this ability. They transplanted testicular cells into fish embryos (pictured) and found that some of these cells could indeed generate fertile eggs in females, as well as sperm in males.



R. PICKETT/ECOSCIENCE

GENETICS**Pregnancy problems***Nature Genet.* doi:10.1038/ng1740 (2006)

Researchers in Canada have identified a gene that could shed light on why some pregnancies result in miscarriage or stillbirth.

An international team headed by Rima Slim at McGill University Health Centre in Montreal studied families in which women had repeatedly suffered from a rare problem known as a hydatidiform mole. This causes the embryo to develop abnormally. In pregnancies in which the hydatidiform mole was absent, the women also had other problems such as miscarriages.

The team says a mutation in the gene *NALP7*, which may affect a mother's immune response to her embryo, could be responsible. They also suggest that defects in the processes that this gene regulates could contribute to pregnancy problems in other women.

PHYSICS**On the scales***Phys. Rev. Lett.* **96**, 042504 (2006)

Weighing an atomic nucleus can reveal how tightly the constituent protons and neutrons are stuck together. According to Einstein's energy-mass equivalence, the binding energy is equal to the difference between the nuclear mass and that of its isolated particles.

Ari Jokinen of the University of Jyväskylä in Finland and his colleagues took advantage of this principle to record the most accurate ever measurements of unusual isotopes of the elements strontium, zirconium and molybdenum. Their ion-trap experiments

show that the most neutron-rich nuclei have less binding energy than theory predicts. These nuclei are thought to be highly non-spherical, and such deformations may throw the predictions off.

CHEMISTRY**Shape-shifting surprise***Angew. Chem. Int. Edn Engl.*

doi:10.1002/anie.200502787 (2006)

Chemists hoping to make novel semiconductors with catalytic properties have been surprised by a molecular reorganization.

Crystal lattices with large channels running through their structure are useful as molecular sponges and catalysts. Mercuri Kanatzidis and Nan Ding of Michigan State University, East Lansing, wanted to introduce such channels into selenium-based semiconductors, opening up the potential for electrochemistry inside the structures.

But treating cubic crystals of the selenium compound $K_6Cd_4Sn_3Se_{13}$ with hydroiodic

acid, which the researchers hoped might make the substance more porous, prompted the compound to form tetrahedral crystals instead (internal structure pictured below).

The crystals allow ions to be readily displaced by other metals, so the new compound might be useful for absorbing traces of heavy metals, the pair suggests.

DRUG DEVELOPMENT**Antibiotic fights back***J. Am. Chem. Soc.* doi:10.1021/ja0572912 (2006)

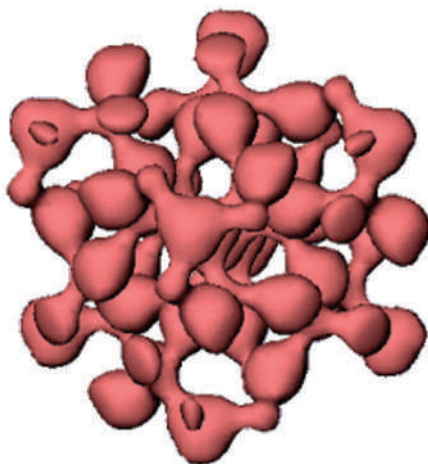
Chemists have synthesized a version of the antibiotic vancomycin that can kill bacteria that are resistant to the drug.

Vancomycin is one of the last lines of antibiotic defence, but some bacteria shrug it off. A single atom of one cell-wall molecule is altered in these strains, limiting the antibiotic's ability to bind to the bacteria.

It took Brendan Crowley and Dale Boger of the Scripps Research Institute in La Jolla, California, 24 chemical steps to manufacture a form of vancomycin with a compensating one-atom change. The molecule can kill both vancomycin-susceptible and vancomycin-resistant bacteria, and could replace the current antibiotic if cheaper manufacturing methods are found.

IMMUNOLOGY**Signal failure***J. Clin. Invest.* doi:10.1172/JCI26091 (2006)

A protein called $\alpha 4$ -integrin, which is found on the outer surface of cells, contributes to the immune-system malfunction that underlies conditions such as multiple sclerosis. But targeting the protein directly



N. DING & M. G. KANATZIDIS

causes severe side effects, as it is also involved in the development of immune cells.

One alternative approach might be to disrupt the signalling pathway that triggers the autoimmune reaction. Mark Ginsberg's group at the University of California, San Diego, has developed mice that carry mutated $\alpha 4$ -integrin, and suggests that this technique could work.

The mutated protein functions well enough to allow normal immune-system development, but not harmful $\alpha 4$ -integrin signalling. When the mice were treated with a compound that causes inflammation and thus mimics many autoimmune disorders, fewer inflammatory cells were recruited to the affected site than in control mice.

CELL BIOLOGY

A few to a kill

Science doi:10.1126/science.1124514 (2006)
"Surprising and slightly horrific," is how Richard Morimoto of Northwestern University in Evanston, Illinois, describes his finding that just a few extra misfolded proteins can kill an organism.

Morimoto and his team looked at *Caenorhabditis elegans* worms engineered to express strings of polyglutamine protein, which are prone to misfolding. They found that the presence of the strings was enough to prompt seven other proteins, which are normally stable at room temperature, to misfold. This disrupted essential cellular processes and killed the worms.

The finding suggests that a critical mass of abnormal proteins may be a common trigger for the emergence of conditions such as Alzheimer's and prion diseases, which are linked to misfolded proteins.

MATHEMATICS

Physical manifestations

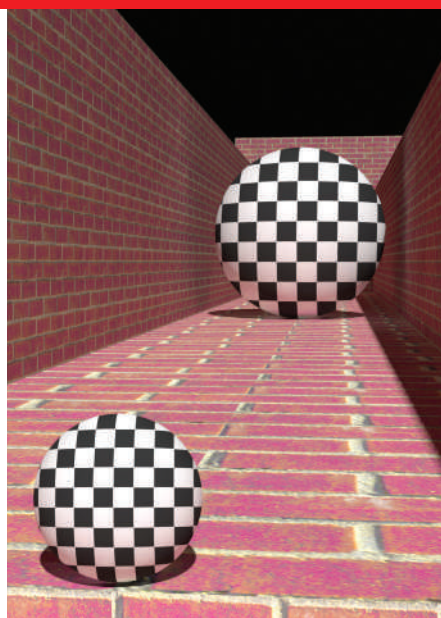
Phys. Rev. Lett. **96**, 040405 (2006)

Forget pens, paper and calculators.

Mathematicians should swap their usual tools for a gas of rubidium atoms if they want to find solutions to an equation known as the Weyl polynomial, suggest Yvan Castin of the Ecole Normale Supérieure in Paris and his colleagues.

Whirlpool-like vortices form in samples of rubidium vapour that have been rotated and cooled to form a quantum-mechanical state known as a Bose gas.

The researchers show that the positions of these vortices can be mapped to the Weyl polynomial, allowing solutions to be obtained by observing the gas.



H. BOYACI/UNIV. MINNESOTA

NEUROSCIENCE

Seeing is believing

Nature Neurosci. doi:10.1038/nn1641 (2006)

The processing of signals in the early stages of the human visual system may be more complex than currently thought.

The first area of the cerebral cortex to process visual signals after they leave the eye's retina is a region known as V1; activity here is believed to depend predominately on physical input, rather than on conscious perception. Scott Murray of the University of Washington in Seattle and his colleagues tested this by taking brain scans of subjects while they viewed three-dimensional computer-generated scenes (pictured above).

Surprisingly, the team finds that activity in V1 depends on the subjects' perception of the size and distance of the object being viewed, not just its image size on the retina. This suggests that theories on the role of V1 may have to be revised.

MICROFLUIDICS

Mini lab makes frugal tests

J. Am. Chem. Soc. doi:10.1021/ja057720w (2006)

Chemists in the United States have developed lab-on-a-chip technology that can screen reaction conditions using less than a microgram of reagents.

Researchers often tweak variables, such as the type of solvent used, to optimize reaction conditions. But this can be labour intensive and wasteful of materials, so Rustem Ismagilov of the University of Chicago, Illinois, and his colleagues developed a microfluidic system to do the job. Tiny drops of reagents are propelled through a tube using a fluorinated fluid. The drops react with a second substance that enters at a T-junction.

The authors suggest the system could also be used to improve the synthesis of drugs and other chemical compounds.

JOURNAL CLUB

Linda Hsieh-Wilson
California Institute of
Technology, Pasadena

**A chemist questions
conventional views on
adding sugar.**

The realization that we have nearly the same number of genes as a worm came as a big surprise. We are taught that DNA encodes RNA, which in turn encodes proteins. So why is it that we are wonderfully complex, and the worm is not?

Hints as to what separates us from simpler organisms may come from studying the modifications made to proteins after they are translated. For instance, the addition of sugars, a process known as glycosylation, is common in higher organisms but generally absent in simple prokaryotes such as bacteria.

In the classical glycosylation model, sugars are assembled and attached in the Golgi and the endoplasmic reticulum, two cytoplasmic organelles. It is commonly believed that, once the glycoproteins reach the cell surface, the sugars will remain unchanged throughout the protein's lifetime. But a recent paper (Q. Chang *et al. Science* **310**, 490–493; 2005) challenges this conventional thinking.

The authors show that a mammalian hormone, klotho, cleaves extracellular sugars on the calcium channel TRPV5, which is a glycosylated protein. This exposes a new sugar residue, leading to activation of the ion channel.

The findings suggest that processing of sugars may be occurring in remarkably dynamic ways, which need not be restricted to the endoplasmic reticulum and Golgi. Over the years, a handful of such non-canonical forms of glycosylation have been identified.

My own research focuses on a dynamic, intracellular form known as O-GlcNAc glycosylation. We've found more than 40 proteins in the brain that are modified in this way, with various functions. What fascinates me about protein glycosylation is nature's ingenious mechanisms for creating diversity.

NEWS

US space scientists rage over axed projects

Proposed cuts to NASA's science budget have unleashed a storm of anger from US astronomers and planetary researchers, who say the reductions would cause irreparable harm and drive young people from the field.

Under a NASA budget unveiled on 6 February (see *Nature* 439, 644; 2006), growth in science spending between 2007 and 2010 would be slashed by 17%. The budget proposed by President George W. Bush has yet to be approved by Congress, but many planned projects — from planet searches to a Mars sample return, as well as scores of individual research grants — are likely to be scrapped (see 'Some cuts proposed at NASA').

Planetary scientist Alan Boss of the Carnegie Institution of Washington says the cuts would devastate US space science — just as physics was jolted when the Superconducting Super Collider was cancelled in 1993, after \$2 billion had been spent on it. "High energy physics never quite recovered from that."

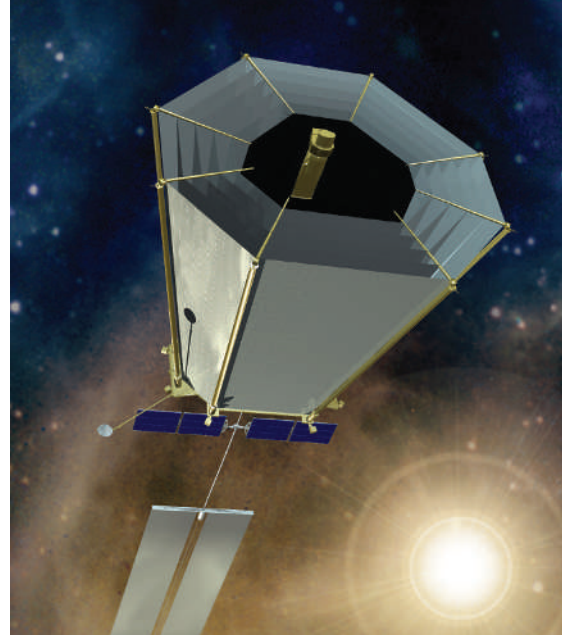
Scientists appreciate that NASA's administrator, Mike Griffin, is struggling to balance his books. Griffin explained during the budget press conference that the science cuts were necessary to pay for shuttle flights required to complete the International Space Station. "It's what we needed to do," he said regretfully.

But Jonathan Lunine, a planetary scientist at the University of Arizona, Tucson, sums up the view of many when he says he finds it "puzzling and frustrating" that NASA would divert money from science, widely considered its most productive enterprise, to keep the aged space shuttles flying. "It seems that NASA is trying to capitalize on its failures rather than its successes," says Lunine.

Particularly hard hit is the search for new planets, a field that appeals to young scientists, says Charles Beichman of the California Institute of Technology, Pasadena. NASA could keep developing technologies for the Terrestrial Planet Finder (TPF) mission given just \$10 million next year, he says.

Instead, the TPF's budget will be wiped out. NASA claims the mission is "deferred indefinitely", says Beichman. "The fact is, they are cancelling the TPF. They are breaking up the technology team."

After budget cuts, this picture of a mission to find earth-like planets is unlikely to become a reality.



There is fury not just at the size of the cuts, but at how they were decided and announced to the science community. Heidi Hammel, a planetary researcher with the Space Science Institute in Boulder, Colorado, says that NASA's advisory council was not operating during much of last year and so "there was absolutely no way to know how these decisions had been made. It's sort of like a black hole over there."

SOME CUTS PROPOSED AT NASA

Space Interferometry Mission to map stars	Delayed by three years
Terrestrial Planet Finder	Deferred indefinitely
LISA, searching for gravitational waves	Deferred indefinitely
Constellation X, a group of X-ray satellites	Deferred indefinitely
Mars Sample Return	Deferred indefinitely/cancelled
NuSTAR, a high-energy X-ray satellite	Will not proceed to development
Europa mission	Not proposed

Disgraced cloner's ally is cleared of misconduct

Gerald Schatten was the Western face of Woo Suk Hwang's stem-cell team, which was recently exposed for faking the results of cloning experiments. On 10 February, Schatten was cleared of misconduct by his university, but chided for taking so much credit for research in which he was barely involved.

The University of Pittsburgh in Philadelphia decided to investigate Schatten in December after claims in a *Science* paper that he had co-authored with Seoul National University's Hwang turned out to be false (see W. S. Hwang *et al. Science*

308, 1777-1783; 2005 and *Nature* 438, 718; 2005). Schatten was senior author on the paper, and his gushing praise of Hwang's research was instrumental in raising the South Korean team's profile in the United States and elsewhere.

The full report has not been released, but in a public summary, the six anonymous investigators conclude that there is no evidence that Schatten knew about the fraud taking place in Hwang's lab, and they applaud Schatten for taking swift action when he became convinced that Hwang's team had

obtained eggs unethically, to create the world's first stem-cell line from a cloned human embryo back in 2004. But they are less pleased with Schatten's decision to name himself senior author on a paper for which his only contribution was editorial.

The summary also points out that Schatten signed a cover letter for the 2005 *Science* paper claiming that all 25 authors of the paper had read and approved of the manuscript, when very few of them probably had. And it notes that Schatten's co-authorship of a 2005 *Nature* paper (see B. C. Lee *et al.*

Nature 436, 641; 2005) reporting the first cloned dog, Snuppy, was based solely on the dubious suggestion "that a professional photographer be engaged so that Snuppy would appear with greater visual appeal". Although stopping short of misconduct, the panel describes Schatten's actions as "research misbehaviour".

Stem-cell researchers contacted by *Nature* generally approve of the report and its conclusions. "Dr Schatten was as much of a victim as the scientific community," says Evan Snyder, who directs the



WINTER OLYMPICS

From the science of artificial snow to extreme medical treatments.

www.nature.com/news

NASA/JPL-CALTECH



Biomedicine to sell itself as a local hero

Lobbyists for biomedical science are changing tactics in an attempt to reverse what they see as a worrying decline in funding. As well as talking generally about the benefits of biomedical research, they plan to tailor their arguments to local economic issues, close to lawmakers' hearts.

In his 6 February budget request, President George W. Bush asked Congress to keep funding flat for the National Institutes of Health (NIH) in 2007. This would mark the fourth consecutive year that NIH funding has not kept pace with inflation, and advocates are worried. "We're going to have to change the way we've done things in the past," says Jon Retzlaff, director of legislative relations for the Federation of American Societies for Experimental Biology (FASEB).

Lobbyists say they aim to dispel the notion that the NIH should be satisfied with the fact that its budget was doubled between 1999 and 2003. They claim that, because of inflation, the agency now has 10% less purchasing power than in 2003, and is on track to issue 3,000 fewer grants in 2007 than in 2003. They also argue that the budget doubling spurred many young people to enter biomedical science. The erosion of that money is leaving these new researchers out in the cold.

"We're eating our seed corn," says FASEB president Bruce Bistran, a molecular biologist at Beth Israel Deaconess Medical Center in Boston, Massachusetts. "We could lose a generation of researchers, or at least several years' worth."

So advocacy groups are going local, by showing lawmakers how NIH funding has benefited their states and home districts. Retzlaff says that FASEB will start with districts served by members of the powerful House budget committee, chaired by Republican Jim Nussle of Iowa. And the Association of American Medical Colleges will emphasize that the University of Pittsburgh Medical Center is the largest private employer in western Pennsylvania.

"There are a lot of places around the country that would like to emulate Pittsburgh," says Dave Moore of the association's office of government relations. "It's important for us to talk about the role the NIH plays as a driver for local economies."

Erika Check

The lack of communication extended even to projects that were being axed. For example, the California Institute of Technology's Fiona Harrison had an Explorer mission that was about to enter its development phase after two years of work. But in what Beichman calls an "egregious breakdown of the process", she learned during the press conference that her NuSTAR X-ray astronomy

satellite had actually been cancelled.

Harrison estimates that about 200 scientists are planning to send petitions or protest letters to NASA. Craig Wheeler, president-elect of the American Astronomical Society, says the society will argue that NASA's science projects should share in the generous increases granted to other research agencies for 2007.

But many space scientists are still just trying

to figure out what it all means — and they believe the draconian cuts won't even fix NASA's larger budget problems. Gregory Junemann is president of the International Federation of Professional and Technical Engineers, NASA's largest union. "Devouring everything else at the agency, while holding out for some future financial miracle, is irresponsible," he says

Tony Reichardt

Y.-J. AHN/AP

Happier times: close ties between Gerald Schatten (right) and Woo Suk Hwang raised the profile of the South Korean's cloning team.



stem-cell programme at the Burnham Institute in La Jolla, California. "Ultimately a collaboration comes down to trust."

Arnold Kriegstein, who directs stem-cell work at the University of California at San Francisco, agrees, but says he is disappointed that

Schatten, as "the first line of defence" against fraud, did not spot problems earlier. "It's hard not to think of Schatten as partly a victim, but on the other hand we were all let down by the lack of careful scrutiny."

However, George Annas, a

bioethics professor at Boston University School of Public Health in Massachusetts, is deeply disappointed by the report, which he calls "pathetic". Signing off a cover letter claiming all the authors approved of the manuscript was clearly wrong, he says. He feels that the race to clear high-profile research hurdles will always bring the temptation to cut corners, and that the report is too easy on Schatten: "The university is basically saying, we will treat you pretty good if you get caught."

Donald Kennedy, editor-in-chief of *Science*, is also uneasy. "The report raises questions," he says. "Nobody I know knows what

'research misbehaviour' is." He adds that *Science's* own review will look further into the legitimacy of Schatten's senior authorship on the 2005 paper. "I thought that he had been over there [to Hwang's lab], and that he was involved with experimental strategies," he says.

Schatten has kept out of the public eye since his break with Hwang in November last year. Like Schatten, officials at the University of Pittsburgh have declined to be interviewed. The full report has been submitted to the dean of Pittsburgh's medical school, Arthur Levine, who will decide whether Schatten should face disciplinary action. **Emma Marris and Erika Check**

Prestige is factored into journal ratings

Journal rankings should measure quality, not just quantity, say researchers who are proposing a new way to assess the status of science publications. Whereas the commonly used impact factor simply measures the number of citations per paper, the researchers say their ranking scheme also measures the significance of those citations, giving a truer measure of a journal's standing in the community.

Ranking journals and publications is not just an academic exercise. Such schemes are increasingly used by funding agencies to assess the research of individuals and departments. They also serve as a guide for librarians choosing which journals to subscribe to. All this puts pressure both on researchers to publish in journals with high rankings and on journal editors to attract papers that will boost their journal's profile.

The most popular index of a journal's status is the ISI Impact Factor (IF), produced by Thomson Scientific. It counts the total number of citations a journal's papers receive, and divides it by the number of papers the journal publishes. But the rise of online journals, coupled with sophisticated search engines that permit rankings of web resources, is triggering a wave of other measures. Last year, for example, physicist Jorge Hirsch of the University of California, San Diego, proposed a metric called the

h-index for assessing the quality of researchers' publications (see *Nature* **436**, 900; 2005).

Now Johan Bollen and his colleagues at the Research Library of Los Alamos National Laboratory in New Mexico are focusing on Google's PageRank (PR) algorithm. The algorithm provides



All for impact: a single metric called the ISI Impact Factor is the most popular tool for ranking journals.

a kind of peer assessment of the value of a web page, by counting not just the number of pages linking to it, but also the number of pages pointing to those links, and so on. So a link from a popular page is given a higher weighting than one from an unpopular page.

The algorithm can be applied to research publications by analysing how many times those who cite a paper are themselves cited. Whereas the IF measures crude 'popularity', PR is a measure of prestige, says Bollen. He predicts that metrics such as the PR ranking may come to be more influential in the perception of a journal's status than the traditional IF. "Web searchers have collectively decided that PageRank helps them separate the wheat from the chaff," he says.

Bollen, however, proposes combining the two metrics. "One can more completely evaluate the status of a journal by comparing and aggregating the different ways it has acquired that status," he says. Some journals, he points out, can have high IFs but low PRs (perhaps indicating a popular but less prestigious journal), and vice versa (for a high-quality but niche publication). Using information from different metrics would also make the rankings harder to manipulate, he adds. So Bollen and his colleagues propose ranking journals according to the product of the IF and PR, a measure they call the Y-factor.

Whereas the top ten list by IF includes many journals that publish only review articles, or that serve primarily as data resources, the Y-factor ranking pushes up journals widely regarded as publishing prestigious original

Plans to pare down climate centre anger UK ecologists

Britain's ability to respond to the threats of climate change and pollution will be damaged if plans to downsize a key research institute go ahead, ecologists have warned.

The Centre for Ecology and Hydrology (CEH) risks losing a third of its 600 staff and half of its research sites if its main funder, the Natural Environment Research Council (NERC), follows through with reform proposals.

The council says the move will leave the centre with a more focused agenda and with critical science projects intact. But many are unconvinced, pointing out that a holistic approach to studying the impacts of climate change is more important than ever.

"We rely on the CEH's papers

when making decisions," says Richard Jefferson, an ecologist at government-funded conservation body English Nature. "It seems ridiculous to be pruning this work back." His organization is one of several to have submitted objections to the proposal under a NERC-run consultation exercise that ended on 15 February.

Shaky finances have left the CEH with an annual deficit of around £1 million (US\$1.7 million) over the past three years. And the government wants the NERC and other research councils to shift resources from dedicated research institutes to competitive grants programmes.

To address these issues, the council wants to reduce the centre's core budget, currently around



B. CARPENTER/CEH

British countryside could suffer because of cut-backs to a key research centre.

£20 million, to £15 million. After consulting with the CEH, the NERC proposes to close its research sites at Banchory, Dorset, Monks Wood

and Oxford, while retaining just four centres at Bangor, Edinburgh, Lancaster and Wallingford.

Once the £45-million costs of the

TOP 10 JOURNALS AS RATED BY DIFFERENT METRICS

Rank	ISI Impact Factor		PageRank ($\times 10^3$)		Y-factor ($\times 10^2$)	
	Value	Journal	Value	Journal	Value	Journal
1	52.28	Annu. Rev. Immunol.	16.78	Nature	51.97	Nature
2	37.65	Annu. Rev. Biochem.	16.39	J. Biol. Chem.	48.78	Science
3	36.83	Physiol. Rev.	16.38	Science	19.84	N. Engl. J. Med.
4	35.04	Nature Rev. Mol. Cell Biol.	14.49	Proc. Natl Acad. Sci. USA	15.34	Cell
5	34.83	N. Engl. J. Med.	8.41	Phys. Rev. Lett.	14.88	Proc. Natl Acad. Sci. USA
6	30.98	Nature	5.76	Cell	10.62	J. Biol. Chem.
7	30.55	Nature Med.	5.70	N. Engl. J. Med.	8.49	JAMA
8	29.78	Science	4.67	J. Am. Chem. Soc.	7.78	Lancet
9	28.18	Nature Immunol.	4.46	J. Immunol.	7.56	Nature Genet.
10	28.17	Rev. Mod. Phys.	4.28	Appl. Phys. Lett.	6.5	Nature Med.

research (see table). For example, among physics journals, the IF places *Reviews of Modern Physics* at the top of the list, but the Y-factor shifts the emphasis to rapid-publication journals. *Physical Review Letters* is the most influential, with a Y-factor of 5.91×10^{-2} . (Declaration of interest: *Nature* receives a very high Y-factor.)

Reinhardt Schuhmann, an editor on *Physical Review Letters*, calls the proposal "an interesting idea", but thinks that such metrics aren't really needed to prove status. "We don't pay much attention to impact factors," he says. But for Bollen, ranking journals more effectively by combining different ranking systems could help protect the integrity of science. He warns that scientists and funding agencies have used the ranking system well beyond its intended purpose. "We've heard

horror stories from colleagues who have been subjected to evaluation by their departments or national funding agencies which they felt were strongly influenced by their personal IF," he says. "Many fear this may eventually reduce

"PageRank helps web searchers separate the wheat from the chaff."

the healthy diversity of viewpoints and research subjects that we would normally hope to find in the scholarly community."

Jim Pringle, vice-president of development at Thomson Scientific, is also keen on the idea. "We have always advocated that research evaluation should be derived not only from metrics such as the IF but also from a thorough knowledge of research content," he says. "Journal status metrics such as this, used in combination with our data, should be encouraged."

Philip Ball

♦ www.arxiv.org/abs/cs.GL/0601030

closures have been met, the NERC will plough the savings into other programmes. But many researchers feel that the CEH's multidisciplinary work cannot be replicated elsewhere.

The centre brings together fields such as chemistry and wildlife management to answer broader environmental questions. Such an approach, many policy experts say, is the best way to generate predictions on issues such as climate-change impacts that can be used by politicians.

For example, at the CEH site in Wallingford, west of London, researchers work with climate scientists and ecologists to model how global vegetation will respond to climate change. "The CEH provides the glue between the climate and

vegetation models," says Ian Woodward, an ecologist at the University of Sheffield who has worked with Wallingford researchers. "It's absolutely critical."

Although the site in Wallingford is not slated for closure, cuts will also be made at the retained sites. So researchers there are nervous about their future.

Pat Nuttall, director of the CEH, says she is currently consulting with programme directors about possible cuts. She stresses that certain crucial projects, such as collecting long-term data sets on biodiversity and water quality, will be retained.

Members of the NERC's

18-strong council, which will meet on 8 March to consider the consultation results, add that the reforms need not cause substantial damage. They argue that reducing duplication of research and

"This is not a carving back on science."

high infrastructure costs will allow all

crucial CEH work to continue.

"This is not a carving back on science", says council member Sara Parkin, a programme director at Forum for the Future, a sustainable-development charity in London, who backs the reform plan. "I've been an environmental campaigner for 40 years. I know that we need quality evidence."

Jim Giles

Doubts hang over source of bird flu spread

The H5N1 avian flu virus has spread to Africa and the European Union for the first time. Attention is focused on controlling the outbreaks, especially in Nigeria, where, as *Nature* went to press, four farms in two adjacent provinces were confirmed as being affected, with further spread suspected. But experts are also urgently trying to confirm the cause of the virus's geographical spread. Although migratory birds have been widely blamed, some believe that the risks posed by the poultry trade are being overlooked.

Until last summer, the virus was restricted to hotspots in a handful of countries in southeast Asia. But since then it has fanned out across the globe, crossing Eurasia into Europe and Africa. The disease's appearance in Africa marks a huge leap in its range, and opens a new front that could vastly increase the bird reservoir that could spark a human pandemic (see *Nature* **437**, 1212–1213; 2005). As in southeast Asia, rural communities often depend heavily on poultry farming, and live in close contact with both domestic and migratory birds.

The outbreak began in Jaji village, in Kaduna state in northern Nigeria, on 10 January and killed 40,000 battery chickens, although it was reported to the World Organization for Animal Health only on 6 February. Over the weekend, outbreaks in wild swans were reported for the first time in Greece, Italy, Bulgaria and Slovenia. Swans are often sentinels of the presence of wider infections, as their large size means their deaths are the first to be noticed.

Most public-health efforts are focused on preventing further spread of the virus in Africa, and urging local populations to avoid butchering and other practices that could lead to human cases. But carrying out an effective assessment of which species pose the highest risk and which areas could be hit next, as well as setting up appropriate monitoring and control measures, depends on knowing how the virus is spreading. And on this issue experts are divided.

Many blame migrating birds. The gene sequence of the virus in Nigeria is identical to that found in Turkey last summer, and the country is a major flyway for several migratory birds from Eurasia. One large genetic study

"Continued focus on wild birds is misguided and is channelling away valuable resources."

published this month also found that H5N1 was present in otherwise healthy migratory birds in overwintering sites in southern China just before their migration (H. Chen *et al. Proc. Natl Acad. Sci. USA* **103**, 2845–2850; 2006).

But it is difficult to be sure whether migratory birds are sparking new outbreaks in poultry, or whether they pick up the virus from poultry infected by other routes. The same study, which analysed more than 13,000 virus samples, also found that trade and poultry movements were by far the biggest factor driving both the repeated spread of the virus outwards from its cradle in

southern China and its genetic diversity. "The link [to migratory birds] seems to remain circumstantial and speculative," says Nick Davidson, deputy secretary-general of the Swiss-based Ramsar Convention on Wetlands.

"No clear evidence is available," agrees Salim Javed, an ornithologist at the Environmental Research and Wildlife Development Agency in Abu Dhabi. He adds that if migratory birds had played a major role he would have expected widespread outbreaks across India, the Middle East and Africa during the autumn migration. This did not happen.

Outbreaks may have occurred in other African countries but not been reported, says Juan Lubroth, a senior official at the United Nations' Food and Agriculture Organization.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Efforts to control avian flu in Africa may be hampered by uncertainties about its route of transmission.

"There may be points in between that we don't know about, that might connect the dots," he says. Indeed, international agencies are tracking rumours of outbreaks in several African countries, including Mali, Egypt, Malawi and Libya. But Lubroth agrees that "it is too easy to point at wild birds". In many African countries there is scant surveillance of travellers carrying eggs and other poultry products, he points out.

Richard Thomas, a spokesman for Birdlife International, says he is "increasingly concerned that the continued focus on wild birds as the principal means of transmission of the virus is both misguided and a dangerous channelling away of valuable resources". Countries with strict bans on poultry imports from neighbouring infected countries, such as Myanmar, South Korea and Malaysia, are currently free of the disease.

There are previous cases of smuggling and the legal poultry trade causing long-distance spread of H5N1, adds Thomas. He says a 2004 poultry outbreak in Lhasa, Tibet, was traced to the transport of birds from Lanzhou in China's Gansu province, 1,500 kilometres away. "The death of poultry from an identical viral strain in locations many kilometres apart says nothing about how the virus was transferred between the two places." Major railway lines link southern China to Kazakhstan, Russia and Eastern Europe, and many believe these could be a route for transmission of the virus from Asia to Eurasia. ■

Declan Butler

P. UTO MI EKPE/AFR/GETTY IMAGES

Thai dogs carry bird-flu virus, but will they spread it?

Large numbers of domestic dogs and cats in Thailand may be infected with the H5N1 strain of avian flu, *Nature* has learned. Experts are struggling to work out whether such carnivores might be spreading the disease.

In an unpublished study carried out last year by the National Institute of Animal Health in Bangkok, researchers led by virologist Sudarat Damrongwatanapokin tested 629 village dogs and 111 cats in the Suphan Buri district of central Thailand. Out of these, 160 dogs and 8 cats had antibodies to H5N1, indicating that they were infected with the virus or had been infected in the past. "That's a lot," says Albert Osterhaus, a virologist at the Erasmus University in Rotterdam, the Netherlands. "This is definitely something to look into." So far, researchers at Bangkok's Chulalongkorn University have isolated the virus from at least one of the dogs.

Wild cats, including tigers,

are known to be susceptible to the virus, but this is the first scientific study to find it in dogs, suggesting that infection could be widespread. Osterhaus is pressing officials at the UN Food and Agricultural Organization (FAO) and the World Organisation for Animal Health to monitor dogs, cats and other carnivores for H5N1. "It's a gap in our surveillance," he says. "Basically all carnivores seem susceptible."

This study is the first to look at the prevalence of the virus in dogs or cats in the field — despite anecdotal reports of cat deaths near poultry outbreaks. But Osterhaus's team has done experiments showing that domestic cats get ill and die from H5N1, and can transmit the disease to other cats (T. Kuiken *et al. Science* **306**, 241; 2004).

And last month, a team from his lab published experiments showing that infected cats excrete virus in their faeces as well as in coughed-out droplets, suggesting that they

could spread the disease (G. F. Rimmelzwaan *et al. Am. J. Pathol.* **168**, 176–183; 2006).

"It is still uncertain what role, if any, this might play in transmission," says Maria Cheng, a spokeswoman for the World Health Organization. "We do not have a full understanding of the viral load needed for human infection, and whether or not the infection of animals other than poultry might contribute."

Juan Lubroth, a senior FAO animal-health officer, says that poultry — not cats and dogs — is the priority for the agency, whose mandate is food security and agricultural livestock issues. The FAO "has only limited resources", he says.

As there is no evidence of dogs becoming ill or spreading H5N1, Lubroth believes the risk they pose is low. "It may be important, but today I am not able to cover all bases. Tomorrow I could be criticized for being wrong; I will have to accept that responsibility." ■

Declan Butler

Dog's life: pups in Thai villages show that canines can catch avian flu.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

PORNCHAI KITTIWONGSAKUL/AFP/GETTY IMAGES

ON THE RECORD

"It makes me mad that I could have had hot fudge sundaes all these years."

A California woman who has been on a low-fat diet for 30 years responds to a major study revealing that such diets do not reduce the risk of cancer or heart disease.

"It's really sad the girls are winning. This isn't the game they should be winning at."

A paediatrician is dismayed by statistics showing that more US teenage girls than boys are smoking and abusing prescription drugs.

Sources: San Francisco Chronicle, Washington Post

SCORECARD



Science 'journalism'
Petroleum geologists give author Michael

Crichton a journalism award for his novel *State of Fear*, which dismisses the threat of global warming as an unrealistic scenario overhyped by wayward scientists.



Nuclear waste

A Nevada family is threatened after their daughter, a contestant in this year's Miss America pageant, made comments in favour of a planned nuclear-waste repository at Yucca Mountain.

NUMBER CRUNCH

Even as Japan plans to significantly increase the number of whales it kills for scientific research, the Japanese are turning up their noses at the pungent, chewy meat. The government sells on carcasses for public consumption, but no one seems to be wildly enthusiastic.

65% more whale meat reached the market in 2005, compared with a decade earlier.

60% more minke whales will be killed this year compared with last year under government targets.

30% was the drop in the price of whale meat between 1999 and 2004.

Source: Associated Press

Calls to conserve biodiversity hotspots

Last week's announcement that dozens of new species had been found in remote Indonesia has led to fresh calls to protect biodiversity. A team of scientists from Indonesia, Australia and the United States announced their findings on 7 February, and hope that their eye-catching photos of never-before seen plants and animals will put pressure on Indonesia to take care of its natural bounty.

The Foja mountains, an isolated area on the western side of New Guinea, became famous in 1981, when biologist and author Jared Diamond found a long-lost species there, the golden-fronted bowerbird. Scientists funded by the non-profit group Conservation International returned to the region in November. What they discovered as they surveyed 300,000 hectares of pristine forest and 750,000 hectares of nearly untouched land surpassed their expectations.

They found what they believe are twenty new species of frogs, four new butterflies and five new palms. And those who ventured to the top of the mountain in a helicopter found "a bizarre red-faced and wattled honeyeater bird". The male Berlepsch's six-wired bird of paradise, never before seen by Western scientists, put on a display for a female that convinced the team that it is a separate species, not just a new subspecies as previously thought.

Finding so much unknown biodiversity is a rare event, which some conservationists say happens only every few decades. Others compare the find to the discovery of tens of frog species in Sri Lanka in 2002 or the discovery of three new mammals in Vietnam in the 1990s.

The main scientific significance is that it may be easier than previously thought for clusters of unique species to arise.

The area used to be an arc of islands, pushed up over time into an isolated mountain range, so scientists were optimistic about finding endemic species there. But Foja is much smaller than other areas where this has occurred, such as the highlands of the Bird's Head Peninsula to the west and the Huon Peninsula to the southeast. "Our study has shown that the 'little' Foja mountains are big enough to constitute a significant area of species endemism," says Bruce Beehler, a Conservation International scientist who led the trip.

Beehler and others want similar investigations to be carried out elsewhere in New Guinea, as well as stretches of the Congo and Amazon basins. "It confirms my view that



B. BEEHLER/CI

Lost and found: the Berlepsch's six-wired bird of paradise has been rediscovered in Indonesia.

there is still much to be discovered and conserved," says Scott Zona, a palm specialist at the Fairchild Tropical Botanic Garden in Coral Gables, Florida. "Just because Google Earth serves up satellite images of the planet, we should not feel that our world is fully explored."

The scientists realize that analysis of the calls, colours, ecology, description and DNA of their specimens will be needed to convince the scientific world of their findings. Beehler says that the birds, butterflies, plants and frogs discovered are for the most part "clear and unambiguous", but that further work will be needed to clarify the taxonomic status of the mammals.

The palms will also require "painstaking comparisons with known species," says Zona. But the results will be worth it, especially if, as suspected, one is a new species of a rare genus (*Pholidocarpus*) only known in Thailand, Malaysia and western Indonesia. "I am hungry for more details," he says.

Another expedition scheduled for the dry season later this year should find even more species, says Yance de Fretes, an Indonesian tropical ecologist at Conservation International, as the constant rain in December made it difficult to find mammals and butterflies. "The frog people were very happy," he notes.

But the headlines could attract not just

scientific attention but also poachers. According to Asep Purnama, director of the conservationist group ProFauna Indonesia, the country has a vigorous illegal trade in rare and endangered species — especially orangutans, leaf monkeys and parrots — thought to be worth a staggering US\$900 million every year.

De Fretes admits that tradesmen might be able to rent a helicopter and pay locals to get access. But he says this is all the more reason to report the species to those who could protect them. "It is important to get the information to the decision-makers," he says. "We need to show how important it is to conserve this area."

For Indonesian scientists such as Johanis Moge, a plant biologist at the Indonesian Institute of Sciences who discovered four new species of palm, the trip was a rare chance to see their own country's bounty. Conservation isn't normally a priority, he says. "Usually Indonesian scientists don't have enough money."

Indonesia does not have a great record of conservation, as massive deforestation has cut into its biodiversity. But this week Indonesia's president Susilo Bambang Yudhoyono asked the team's leaders to present their case, de Fretes says. If the international media frenzy over the 'lost world' doesn't get his attention, the scientists have another card up their sleeves: they have offered to name one of the new species of butterfly after his wife.

David Cyranoski

"Just because Google Earth gives us satellite images, we should not feel that our world is fully explored."

NIH urged to enforce its public-access policy

The US National Institutes of Health should require researchers to participate in its public-access policy rather than keep it voluntary, says the board of regents of the National Library of Medicine.

The policy requests all NIH-funded investigators to submit their peer-reviewed papers to the free PubMed Central database within one year of publication. In part, this is meant to help the public to access the results of federally funded research. But since the policy was implemented in May 2005, fewer than 4% of eligible articles have been added, the board reports.

Lack of awareness or technical difficulties don't seem to be the problem, as the NIH has repeatedly encouraged researchers and publishers to submit their papers. The board expressed its opinion in an 8 February letter to NIH director Elias Zerhouni.

Japan rejects Korean offer to discuss kidnap case

Japan last week rejected a North Korean proposal to have genetics experts from both countries discuss the DNA analysis of some controversial human remains.

The rejection came at the end of bilateral talks in Beijing, during which Japan failed to make progress in finding out what happened to eight citizens kidnapped during the 1970s and 1980s. North Korea says the remains at the centre of the case are those of the woman Megumi Yokota. Japan said DNA tests show the remains belong to someone else, and demanded to know Yokota's whereabouts.

Last year, Tomio Yoshii, the scientist who carried out the tests, admitted that they were not conclusive (see *Nature* 433, 445; 2005). Yoshii, a DNA expert with the Tokyo metropolitan police, has not since responded to requests for confirmation from either journalists or politicians.



Missing: the family of Megumi Yokota call for Japanese sanctions against North Korea

Plight of polar bears might force US action on climate

Prompted by lawsuits from environmental groups, the US government is considering putting polar bears on its list of threatened species. Activists hope such a listing will force the government to take action on climate change.

Most animals protected by the Endangered Species Act are declining in numbers because humans are taking over their habitats. But the main threat to the bears is thought by many to be climate change, as rising temperatures melt the Arctic sea ice under their paws.

Once a species is listed, agencies are prohibited from taking any action that would jeopardize it. Kassie Siegel, director of the climate programme at the environmental



group Center for Biological Diversity, says she takes this to mean that the US government must control emissions of heat-trapping carbon dioxide more effectively. "We believe it is a very solid legal argument," Siegel says.

On 9 February, the US Fish and Wildlife Service opened a 60-day period for public comments on whether to list the bear.

Patent office lets army alter description of weapons

The US Army is deleting what it claims was a mistaken reference to chemical and biological weapons in a patent application.

In 2003, the army was issued a patent on a grenade that it said could be packed with various payloads, including chemical and biological agents (see *Nature* 423, 789; 2003). Critics were alarmed because international treaties and federal law ban the US government from making and using biological and chemical weapons.

At the time, the army said it had no intention of making such weapons, and asked the US patent office to delete the reference to these agents from the patent. Last February, the patent office refused, but it has now relented.

Critics remain concerned. "Changing the patent doesn't change the weapon," says Edward Hammond of the Sunshine Project, an international body that works to stop the use of biological weapons.

US fails to make the grade on tackling ocean issues

The Joint Ocean Commission Initiative has given the United States a D+ on its efforts to combat problems facing the oceans, such as pollution and overfishing.

Suggestions from the US Commission on Ocean Policy and the Pew Oceans Commission prompted the Bush administration in 2004 to release an ocean action plan. The new report card suggests that little progress has been made in six areas from fisheries management to funding and education.

A White House spokeswoman, Michele St Martin, says that she expects higher marks

when future projects, such as reauthorizing a national fisheries conservation act, get under way.

Methane deposits not to blame for sudden warming

New data from a Greenland ice core have dealt what could be a blow to the idea that methane escaping from the seabed can drive rapid climate changes.

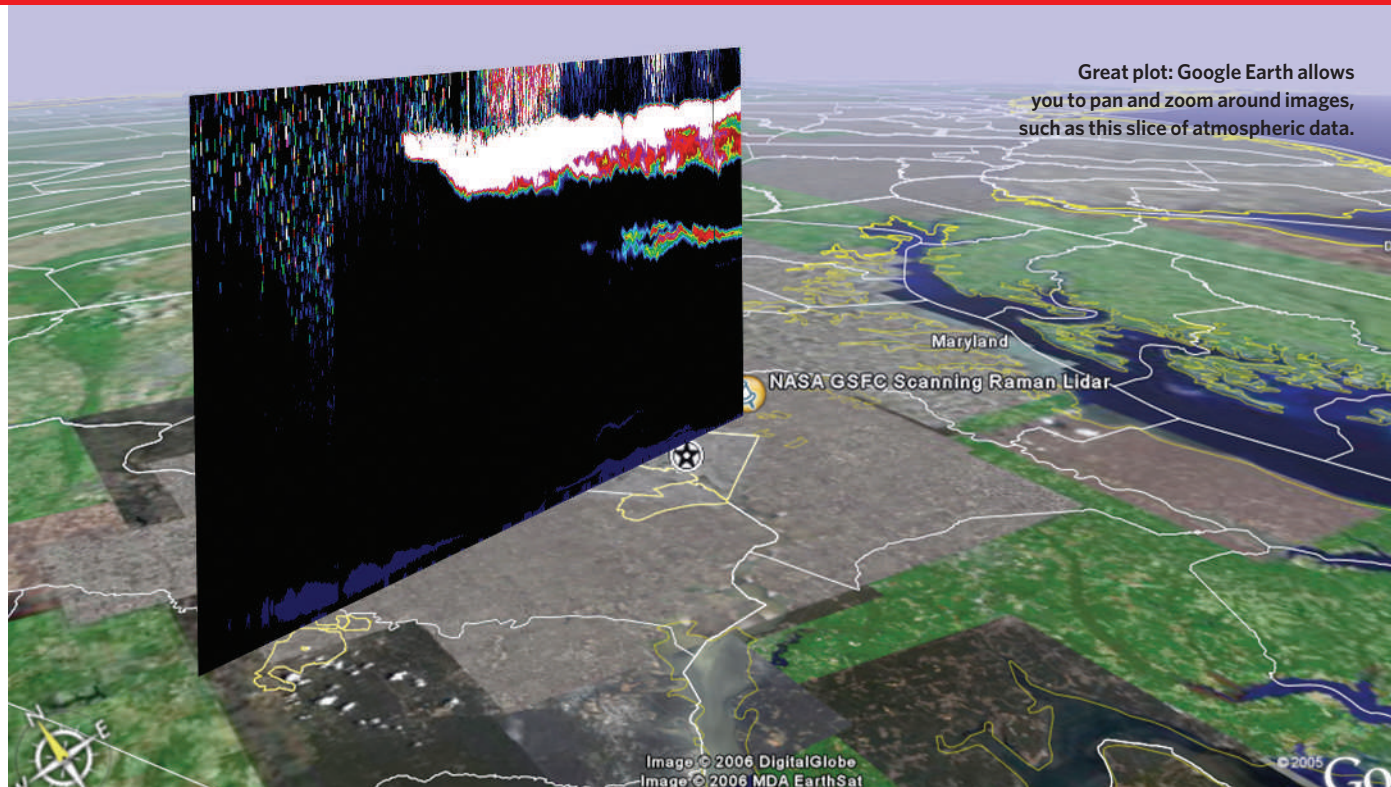
Methane, a strong greenhouse gas, is buried in some sea-floor sediments in a crystal called methane clathrate. A release of trillions of tonnes of methane from clathrates, for example, is thought to have triggered a temperature jump some 55 million years ago at the Palaeocene/Eocene boundary. Other rapid warming events have also been attributed to methane releases.

Now Todd Sowers, an Earth scientist at Pennsylvania State University, has measured the hydrogen isotopes of atmospheric methane from three warming episodes, 38,000, 14,500 and 11,500 years ago. Methane from clathrates contains more of the heavy form of hydrogen than methane from land-based sources. Sowers found no evidence for increased amounts of methane from clathrates in these periods (T. Sowers *Science* 311, 838–840; 2006).

"This means that seafloor methane reservoirs must have been stable at these times, or at least that no significant amounts of methane escaped the ocean," Sowers says.

Correction

In our story "Cloudshine is a stellar snap for Harvard duo" (*Nature* 439, 250; 2006), we incorrectly presented an image from the cited paper as the first picture of cloudshine. Similar images have been taken in the past at optical and other wavelengths.



Great plot: Google Earth allows you to pan and zoom around images, such as this slice of atmospheric data.

ECOTRONICS/GOOGLE EARTH

The web-wide world

Life happens in three dimensions, so why doesn't science? **Declan Butler** discovers that online tools, led by the Google Earth virtual globe, are changing the way we interact with spatial data.

Next month, biologist Erik Born will be wielding a crossbow and firing satellite tags into the hides of walrus, having manoeuvred his rubber dinghy through the pack ice off western Greenland. By tagging the walrus, Born will be able to track the animals' movements and behaviour from afar over several years. He will keep an eye on them using the same free Internet tool that has opened the eyes of millions to the possibilities of digital geography (and the sight of their house from above) — the Google Earth virtual globe.

When the walrus migrate in the spring, Born and anyone else with a copy of the Google Earth software and a decent Internet connection, will be able to follow their westerly path to Baffin Island or the Canadian coast, and their return.

Born, who works at the Greenland Institute of Natural Resources in Nuuk, got the idea from his colleague Leif Toudal Pedersen, a remote-sensing researcher at the Technical University of Denmark in Lyngby. Last month, Pedersen began using Google Earth to visualize live data from satellites, recording the density and drift of Arctic ice, as well as the position of individual buoys and icebergs. Born's decision to follow suit means they can collaborate easily, with a tool that is free and convenient.

Combining Born's tracking data with Pedersen's maps should reveal how changes in ice affect the walrus' movements and behaviour.

With traditional Geographic Information Systems (GIS) software — which was previously the only way to deal with spatial data like these — combining the two data streams would have been a headache. With Google Earth it will be effortless, says Pedersen: "It provides a very easy interface to a lot of different data."

Home page

To the casual user, of which it has attracted millions since its launch last June, the appeal of Google Earth is the ease with which you can zoom from space right down to street level, with images that in some places are sharp enough to show individual people. Its popularity with a growing number of scientists lies in the almost-equal ease with which it lets them lay data with a spatial component on top of background imagery — a trick they can repeat with multiple data sets. By offering researchers an easy way into GIS software, Google Earth and other virtual globes are set to go beyond representing the world, and start changing it.

Leaving aside its value as a tool for meshing data, Google Earth is simply an excellent visualization aid. It really comes into its own when you 'tilt' the browser away from the default vertical view, and fly along in three dimensions, swooping over mountains and through valleys, or along city streets with buildings looming up on all sides.

David Whiteman, an atmospheric scientist at NASA's Goddard Space Flight Center, is using this fly-by feature to understand local weather systems. Meteorological phenomena that occur over distances between 2 and 200 km are poorly predicted by current weather models, and scientists badly need real-time observations to refine their predictions.

So this summer, Whiteman will take part in a large field project, observing a three-dimensional column of the atmosphere centred on the Howard University campus in Beltsville, Maryland. Measurements will be made simultaneously by a battery of aircraft and satellite-borne instruments, including Whiteman's Raman Airborne Spectroscopic Lidar, which measures the properties of atmospheric water vapour, aerosol particles and cloud droplets.

This exercise will generate masses of data on the formation and behaviour of clouds, which the three-dimensional capabilities of virtual globes are ideally suited to viewing. "We are going to be able to display curtains of data of different atmospheric quantities, looking into them from above and flying along them," explains Whiteman. "We will have the topology of the terrain; you will be able to see uplift and cloud formation, and its interaction with precipitation." And like Born's work, these easily viewed data can be shared instantly with colleagues.

"I think it's just a short matter of time before

systems like Google Earth are an essential requirement for people in our field," says Whiteman. "As soon as one group shows that this is useful, everyone will adopt it."

Indeed, increasing amounts of scientific data are becoming available, often in real time, in formats that can be displayed by virtual globes. Millions of species distributions from the Global Biodiversity Information Facility, headquartered in Copenhagen, Denmark, are available for Google Earth, for example.

Digital watch

The eventual impact will be nothing less than the realization of a 'digital Earth', as described by former US vice-president Al Gore in 1998. "Digital Earth was always intended to allow us to 'fly' from space (virtually, of course) down through progressively higher resolution data sets to hover above any point on the Earth's surface — and then display information relevant to that location from an infinite number of sources," says Gore. "Its highest purpose was to use the Earth itself as an organizing metaphor for digital information." But the project died a death in 2001 after Gore lost the 2000 US presidential election.

Michael Goodchild, a GIS expert at the University of California, Santa Barbara, says that scientists' use of virtual globes is breathing fresh life into Gore's dream. There is renewed hope that every sort of information on the state of the planet, from levels of toxic chemicals to the incidence of diseases, will become available to all with a few moves of the mouse.

The GIS software is already an important tool for understanding spatial and temporal factors in a wide range of disciplines. But many

scientists that could use GIS do not, and it has remained largely the preserve of specialists. Goodchild is convinced that tools like Google Earth will increase awareness of GIS's potential and encourage researchers to explore more powerful GIS techniques. "It's like the effect of the personal computer in the 1970s, where previously there was quite an élite population of computer users," Goodchild enthuses. "Just as the PC democratized computing, so systems like Google Earth will democratize GIS."

Goodchild illustrates the system's ease with an anecdote from his own teaching. "Typically, I used to spend an entire year taking senior undergraduates through courses in GIS. And at the end of the year, as a treat, I might let them generate a three-dimensional fly-by over a landscape," he says. "Now, using Google Earth, a ten-year old can do that." Goodchild thinks that once scientists experience this easy visualization, they will be drawn into deeper forms of analysis using the powerful techniques that GIS professionals have developed over decades.

GIS techniques use statistics to predict or explain patterns, whether they be outbreaks of vector-borne disease or urban crime waves. Displaying such patterns on the current version of Google Earth requires gentle hacking. But Brian McClendon, director of engineering at Google Earth, says he wants as many people as possible to use the program — and will consider adding features that make it easier for them to get their data into it. McClendon is hoping to attract scientists who will generate interesting content for Google Earth's other customers.

Whiteman emphasizes, however, that three-dimensional visualization is about more than



E.W. BORN

The time has come: online tools will revolutionize the way walruses and other animals are tracked.

just creating images. He says it is critical for generating scientific hypotheses and the questions they go on to test. "What have I measured; what are the relationships I have here; what can I explore? It is part of getting an intuitive grasp of the problem, the measurements and the analytical challenge."

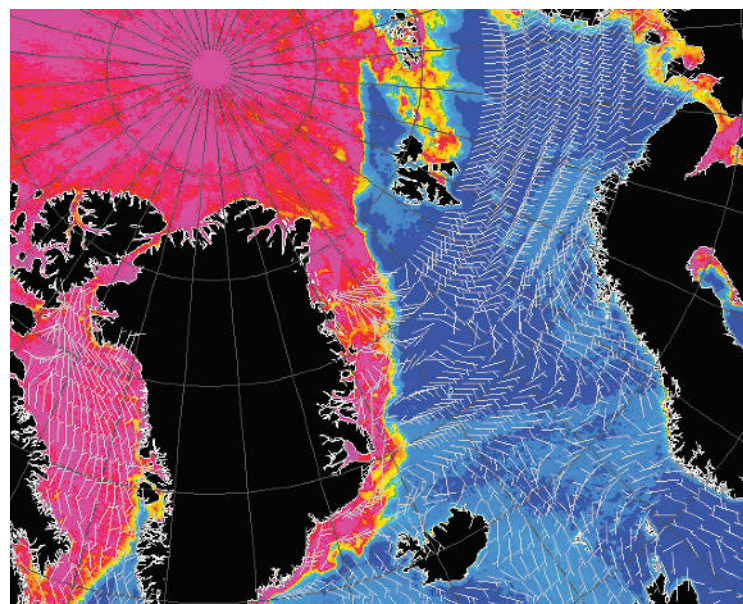
If one can simultaneously visualize the data and the predictions from models, areas where the data agree or disagree with the model jump out, he says. "This could dramatically improve the efficiency of exploring relationships among quantitative data."

Popular movement

Google Earth has no analytic functions and is not designed to replace professional GIS software; in fact, it should be a boon to the software makers. "Google Earth is just the most fantastic thing I have ever seen," says Jack Dangermond, founder and president of ESRI, the world's largest creator of GIS software. ESRI, which is based in Redlands, Virginia, and other GIS companies that were caught napping by Google Earth are now eager to capitalize on its success. They are creating a slew of products that combine its ease of use with their traditional analytic strengths. "It really is opening up our world," says Dangermond, "and business is booming."

Later this year, ESRI will release a huge upgrade of its flagship desktop GIS product, ArcGIS, which will let users publish virtual globes on the Internet and analyse their data in many ways. "We have taken several thousand functions and deployed them in an Internet environment," says Dangermond.

As part of the package, ESRI will also release a free visualization tool, ArcGis Explorer, which some GIS professionals are calling a Google-Earth killer. Data in Google Earth need to be in a specific format; ESRI's tool will allow users to view not only data from ESRI's



Pretty useful: virtual globes can display live data such as sea-ice distributions and wind directions on the same map.

L.T. PEDERSEN/TECHNICAL UNIV. DENMARK

own products, but also information in formats that are being increasingly standardized through the Open Geospatial Consortium. This international body is working to ensure that computers can understand descriptions of the spatial features of anything from highways and postcodes to icebergs.

Unlike Google Earth, the ESRI viewer comes equipped with a series of analytic tools. Scientists can run models on their servers, and simultaneously view them over the Internet in ArcGis Explorer by dragging and dropping data files. They can fuse multiple data sources on screen, and export them in whatever format they choose. Still, McClendon believes that both Google and ESRI will profit from a generally increased interest in GIS, and says he welcomes the competition. "I think that we want many more people to understand and care about GIS, and ESRI's tools are the best in the business for that," says McClendon.

Skyline Software Systems, based in Chantilly, Virginia, was one of the first companies to offer a virtual globe: TerraExplorer. Later this year it will release Skyline Online, a browser-based tool similar to ArcGis Explorer. "It will empower users at home as well as researchers with capabilities that have been available only to government agencies until now," says company president, Ronnie Yaron.

Globe trotters

Meanwhile, in the public sector, NASA has no intention of using its World Wind virtual globe to compete with Google, at least according to one scientist on the project, who did not wish to be identified. He points out that World Wind is explicitly designed for scientific information, and its code is open source so that scientists and software developers can tailor it to their needs. He admits that World Wind's performance lags behind that of Google Earth: it is slower and eats up much more memory and processor time. But NASA intends to remedy these performance issues — a major upgrade will be released in July, he says.

Joshua Been, GIS librarian at the University of Texas, Arlington, sees firsthand how Google Earth is promoting interest in a range of professional GIS systems. Researchers and students increasingly come to him wanting to use Google Earth or World Wind, but most of their queries need analyses beyond straightforward visualization, he says. "So I try my hardest to show them the super-cool capabilities of full-blown GIS."

Still, Been's library offers the same level of support for the two free systems as it does for professional software, and he encourages people to use them. "If they want GIS data to be



Free ride: creators of geospatial software such as Jack Dangermond say that "business is booming".

viewable in a free and public three-dimensional viewer, Google Earth will continue to be the standard," he says. "I disagree entirely that ArcGIS Explorer will be a 'Google-Earth killer.'"

For Ming-Hsiang Tsou, a geographer at San Diego State University in California, "the year 2005 was a watershed for GIS on the Internet". He notes in particular how natural disasters, such as the tsunami and Hurricane Katrina, etched the power and utility of GIS and Google Earth into people's minds (see page 787). Before Google Earth appeared, Tsou

"We are going to be able to display curtains of atmospheric data and look into them from above or fly alongside them."

— David Whiteman

himself created interactive two-dimensional Internet maps of the 2003 San Diego wildfires, featuring data on fire spread, hot spots and rescue operations.

The huge public interest in such visualizations means scientific information is being made much more accessible to users, he says. Meteorological radar data and satellite images have long been available online, but Google Earth is allowing the agencies that provide them, such as the US National Oceanic and Atmospheric Administration (NOAA), to make the data more useful and user friendly.

Real-time weather information can now be displayed in Google Earth alongside the landmarks and routes in which the general public is interested, says Valliappa Lakshmanan, a NOAA researcher at the University of

Oklahoma, Norman. People can use Google Earth to ask: "How far is the rain core from the route that I will take this afternoon? Is it near grandma's house?" Such detail is possible because the resolution of forecast data is now as good as 1 km, updated every 120 seconds.

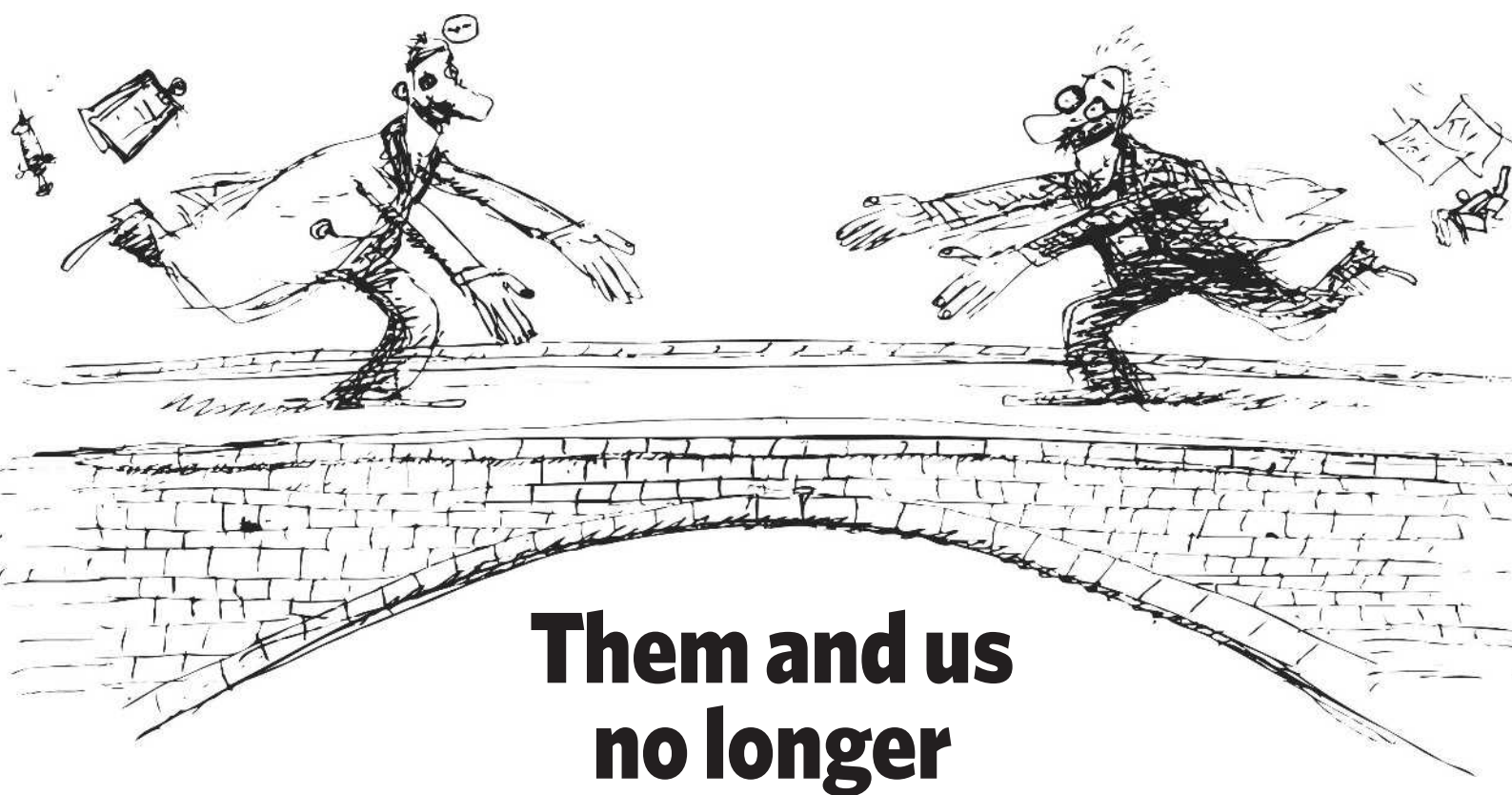
The value of scientific data increases the more you can link it to information that the user already considers important, says Lakshmanan. "Scientists should take this opportunity to use GIS to present their scientific results in a way that users can easily tie to other data sources."

Nowhere is this more true than in the environmental sciences. One of the traditional roles of GIS has been to provide data to support decision-making. And environmental groups that have discovered GIS are starting to use it to change the balance of power in public debates. As more citizens become concerned about their local environment, easy-to-use virtual globes will facilitate the communication of spatial information between stakeholders and government agencies.

Back in Greenland, Born and Pedersen hope to be able to record the effects of climate change on the spread of sea ice and on their tagged walrus. If the ice continues to recede in the Davis Strait, as it has for the past two decades, they will be able to monitor its immediate impact on walrus feeding grounds. Is it too much to hope that, in future, the same approach to sharing data will help us decide what to do about global climate change? ■

Declan Butler is a senior reporter for Nature based in Paris.

For more on Google Earth, see page 763 and www.nature.com/news.



Them and us no longer

Scientists and medical doctors view research through different lenses — but the gulf in outlook between the two tribes isn't what it used to be. **Meredith Wadman** reports.

Genetics graduate student Rima Adler reached the limits of her collegiality one evening last May, when she needlessly missed a concert by the Baltimore Symphony Orchestra.

Late that afternoon, she had stayed in the lab to help a physician colleague, and ended up spending three hours explaining the basics of the polymerase chain reaction and how it could help the young physician's analysis of their results.

Adler ended up stuck in the lab until 9.30 at night, forgetting about her concert ticket. So she wasn't too thrilled the next day, when she heard the same colleague asking someone else in her lab at the National Institutes of Health (NIH) in Bethesda, Maryland, the same question. "I thought: what the hell am I wasting my time for?" Adler recalls. She thinks the colleague "didn't want to learn new things". It was a classic case, she says, of a physician "wanting to just get an answer" instead of taking time to properly understand a problem.

Conflict between PhDs and MDs has been as perennial in biomedical research labs as head lice are in kindergarten — and can be just as unpleasant. "There is this stereotypical idea among medical students that those people are getting PhDs because they couldn't get into med school," says Javed Siddiqi, a California-based neurosurgeon who spent two years in a research lab before himself going to medical

school. Meanwhile, he says, PhD students see themselves as intellectuals, "broadening the frontiers of science" while medics, as he puts it, are "memorizing anatomy".

The culture clash arises when freshly trained doctors are unleashed on laboratories full of bench scientists. Having been deferred to by patients and taught to treat and discharge people in three days — or even three hours — they encounter a strange environment where asking questions is as important as obtaining answers, quick fixes are rare, and jeans and trainers trump ties and stockings. And lurking always in the background is the fact that the MDs are paid higher salaries.

"The cliché — which must have some truth in it — is that the medics want to come into the lab and get everything to work very quickly, without any special struggle, and then write some fantastic paper and make their career in five minutes," says Paul Klennerman, a virologist with an MD and PhD at the University of Oxford, UK. "They generally get a bit of a shock when they realize quite how much effort is involved and how much can go wrong."

But Peiman Hematti, a physician with extensive bench training who studies stem cells at the medical school of the University of Wisconsin-Madison, says the problem has diminished over the years. He recalls a conversation he had ten years ago when he was a young MD looking to get training as a

postdoc: a friend cautioned him in the direst terms about the inhospitable environment he was sure to find if he joined a lab headed by a scientist rather than a physician.

Conflicts happen

In fact, Hematti's experience suggests that the stereotypes are losing their power. Conflicts "can happen in any lab", he says. He doesn't think they occur because physicians expect scientists to clean up their messy lab benches for them, or because PhDs look down their noses at MDs. Today "it's more a personality thing; I don't think it matters whether there's an MD or a PhD at the end of your name."

During a postdoctoral fellowship, for example, Hematti enjoyed what he terms a "very fruitful" collaboration with Boris Calmels, a French scientist whose strengths in molecular biology combined with Hematti's clinical background to produce notable papers^{1,2} on the integration of lentiviral and retroviral vectors into host genomes. "The result was much better than if we had worked separately," he says.

Calmels, now at the Centre for Cell and Gene Therapy at the Paoli-Calmettes Institute in Marseille, France, agrees. "When he had basic questions, usually I had the answer," he recalls. "And when I was looking for specific clinical knowledge, he was the person to ask."

A number of forces have combined to blunt the once-common tensions between scientists

and physicians in the lab. For a start, clinical medicine is less hierarchical than it used to be, and produces doctors less likely to enter the lab with a “Scalpel, nurse!” attitude. And in the United States, at least, sharp growth in biomedical-research funding, combined with growing demands for the translation of research into cures, is blurring the line between bench and bedside.

Birth of a divide

In this environment, cooperation has become imperative. Isolated efforts by physicians or bench scientists “just don’t work”, says Rick Morgan, a geneticist who oversees neurosurgeons’ postdoc work at the National Cancer Institute in Bethesda, Maryland.

Others argue that tensions have never been as great as laboratory gossip would suggest — and were always far outweighed by the achievements of MD–PhD partnerships. These stretch back to Linus Pauling and Harvey Itano’s classic 1949 discovery that sickle-cell anaemia is caused by a structural problem with the haemoglobin protein found in red blood cells. “The mix of MDs and PhDs in my lab enriched everybody’s experience,” says Sam Silverstein, a physician in the department of physiology and cellular biophysics at Columbia University in New York.

Calmels, however, suspects that the problem is more acute in nations such as his own, where the two groups are educated separately from the moment they leave school. Moreover, in France, training rules may force physicians seeking academic appointments to complete a year of bench research. “They’re not really interested in working at the bench,” Calmels says. “Most of the time, there are two different cultures.”

One hundred years ago, this division would have seemed odd. Biochemistry was then the nexus of medical research, but there was no such thing as a PhD in biochemistry: aspiring

biochemists went to medical school on their way to the bench. This explains why the biomedical luminaries of the first half of the twentieth century, from Otto Meyerhof to Hans Krebs, started out as medical doctors. And although they spent the rest of their lives in research, their experience in clinical medicine provided an important bridge between the two.

But after the end of the Second World War and a sudden influx of public funds, medical schools for the first time established academic departments with PhD-granting powers. Ironically, this meant that PhD students were suddenly being trained on the premises of major

“People are realizing that their specialty is not an island; they are going to have to talk to somebody else.” — Javed Siddiqi

medical schools and centres — but with virtually no exposure to patients, doctors or pathology. Thus the great divide was born.

“For many years, it was the prevailing view among scientists that somehow medicine was a bit dirty, because people made a lot of money and it wasn’t based on hard science,” notes Irwin Arias, a physician and cell biologist who oversees a popular NIH lecture series that teaches clinical medicine to scientists. “That created this tension. There was ‘us’ and there was ‘them’. The physician was the one who dealt with the patients.”

It wasn’t until the early 1990s that the public push to bring research to the clinic gained momentum, and a slow rapprochement of the two sides began. “The wheel has turned,” says Arias. “More and more there’s the realization that you can’t separate basic science from human health. Those who try to do it, even as scientists, often get butchered” by grant reviewers, he says.

Meanwhile, those paying for research are moving to address the issue directly. This week, the Howard Hughes Medical Institute (HHMI) is unveiling awardees in a new \$10-million grant programme aimed at directing PhD students into clinically related thesis projects co-mentored by a physician and a scientist. Thirteen awards are being made following 82 applications. “It’s an idea whose time may have come,” says William Galey, director of graduate education programmes at the HHMI.

And at the NIH main campus, a weekly lecture series schooling scientists in medical topics from tuberculosis to Parkinson’s disease is drawing enthusiastic reviews. The ‘Demystifying Medicine’ series attracted 80 or so scientists to an NIH auditorium when it was launched five years ago; this year more than 900 have registered for it, and thousands all over the world download videos of the lectures.

“Nowadays, PhDs are intensely interested in learning more about disease — what it looks like, what it feels like,” says Arias, the course designer. “And it turns out that it’s a lot easier to demystify medicine for PhD scientists than it is to make last year’s chief resident into a bench scientist.”

That transition may be the shape of the future, as the era of the physician–scientist draws to a close. Today, “you have to be super-human to do both well,” says neurosurgeon Siddiqi. So getting the best science to the bedside is going to demand collaboration. “Neither the MD nor the PhD is going to be doing the whole thing. More and more, people are realizing that their specialty is not an island; they are going to have to talk to somebody else.” ■

Meredith Wadman is a freelance writer based in Washington DC.

1. Hematti, P. et al. *PLoS Biol.* doi:10.1371/journal.pbio.0020423 (2004).

2. Calmels, B. et al. *Blood* **106**, 2530–2533 (2005).



All aboard: crowded weekly lectures at the National Institutes of Health uncover the mysteries of medicine for scientists.

BUSINESS



Cultural shift: NEC is encouraging a Western style of management at its Chinese laboratory.

China, and many Japanese companies are strengthening their ties with universities in a bid to snag students after they graduate. For example, Osaka-based Matsushita Electric Industrial Company has researchers that teach classes at Dalian University of Technology in the northeastern province of Liaoning. The company, best known for its Panasonic brand, has several labs in China and opened a software development facility in Dalian in 2004.

"It's tough to recruit good students there, so we want them to get interested in us," says Atsushi Ando of Matsushita's research strategy group. Matsushita says that over the next few years it will triple or quadruple the number of researchers and engineers at its Dalian lab from the current 100.

But recruitment isn't the only issue: many Japanese firms are keen to train the Chinese students in their corporate ethos so that their recruits stay put. Hitachi and Fujitsu both say that they are trying to imbue young researchers with Japanese virtues, such as teamwork and corporate loyalty. Since Hitachi established a Chinese lab in 2000, it has invited many of its 60 Chinese researchers to Japan for six months to learn these virtues. "We started from scratch, so educating researchers was the toughest thing for our first five years," says Michiharu Nakamura, a Hitachi vice-president in charge of the Beijing laboratory.

Lab managers may laud the quality of Chinese graduates, but the monthly salaries paid the recruits are only about a quarter of the US\$2,000 or so that similar graduates get in Japan. And managers are aware that US companies such as Microsoft and IBM have a more positive image among students. Many Japanese employers also acknowledge that the research 'culture' in China, with its strong emphasis on individualism and competition, has more in common with that of the United States than with the Japanese model.

This might explain why NEC has let Hsueh take a Western approach to running his lab. His management practices are more Californian than Japanese: English is the working language, and researchers get financial bonuses for filing patents, publishing papers and producing ideas that become incorporated into prototypes. "This approach is a first for NEC," Hsueh says.

Other companies are also shifting responsibility from Japanese managers to Chinese nationals. Many group leaders at Hitachi, for example, are now Chinese. The main challenge over the next few years, Nakamura says, is to develop products there not just for China, but for global markets as well.

Across China's frontier

China has become the preferred place for international firms to open research labs — and Japan is leading the way, reports Ichiko Fuyuno.

Electronics engineer Min-Yu Hsueh wasn't planning to swap California for Beijing, but when the invite came he found it very hard to refuse. A Chinese-American entrepreneur based in San Jose, Hsueh was three years ago working on setting up his next company. Then, out of the blue, Japanese electronics giant NEC offered him the chance to run its first research lab in China.

"They said: 'China is expanding its market, so we want to play there more efficiently, and we need to invent new technologies'," Hsueh recalls. He was impressed by the plans, and now runs NEC Laboratories China in Beijing, leading a team of 40 researchers, most of them computer scientists.

NEC is just one of a host of Japanese companies to have set up research facilities in China in recent years. These firms hope to hire China's well-trained and relatively inexpensive young scientists, and to adapt existing products to the burgeoning Chinese market. The companies are encouraged by generous tax relief for investing in research and development, and hopes that they will be able to secure intellectual property in China (see *Nature* 438, 420–421; 2005).

But for research managers brought up in Japan's system of teamwork and corporate loyalty, the fast-moving, Mandarin-speaking environment takes some getting used to. "It's challenging for them," says Hironori Uchibori, senior economist at Mizuho Research Institute in Tokyo. "But the importance of China is

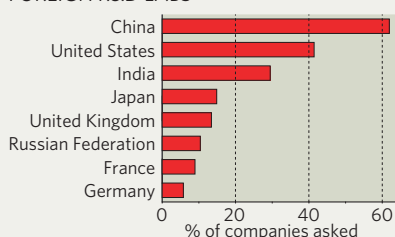
rising, so they are trying to make it work."

In 2005, Mizuho surveyed almost 1,400 medium and large industrial firms in Japan and found that more than 250 had already established research labs in China. Only 116 had done so in the United States.

Japanese firms are not the only ones heading for China. US electronics company Motorola was the first foreign firm to set up a research lab there, back in 1993. Since then there's been a veritable gold rush, and there are now some 700 labs operated by multinational companies, according to a 2005 United Nations report. That report found that firms favoured China more than any other nation as the location for expanding research and development over the next five years (see graph). In the survey, 60% of European companies said that they had plans to expand their research abroad, but fully 90% of Japanese ones said so.

Competition to hire top graduates is stiff in

PREFERRED LOCATIONS FOR FOREIGN R&D LABS



SOURCE: UN

All Correspondence this week arises from recent News and Editorial coverage of fraud and quality control.

Why negatives should be viewed as positives

SIR — Your News story “Journals submit to scrutiny of their peer-review process” (*Nature* 439, 252; 2006) reports findings of no bias against manuscripts presenting negative results. But the journals examined in this study are at the top of their field, and top journals are only likely to receive submissions reporting negative results if these are of clear ‘positive’ interest. This is certainly the case in ecology: Julia Koricheva (*Oikos* 102, 397–401; 2003) showed that non-significant results in ecology tend to be published in lower-impact journals.

This filtering of results undoubtedly biases the information available to scientists (see, for example “Null and void” *Nature* 422, 554–555; 2003). And communication is at the heart of science.

If non-significant results remain unpublished, we will be left with only half the picture. We encourage scientists to submit the negative results of their rigorous research to journals such as the *Journal of Negative Results — Ecology and Evolutionary Biology* (www.jnr-eeb.org), the *Journal of Negative Results in BioMedicine* (www.jnr-bm.com), the *Journal of Negative Observations in Genetic Oncology* (www.path.jhu.edu/NOGO), and the *Journal of Negative Results in Speech and Audio Sciences* (journal.speech.cs.cmu.edu).

Bob O'Hara*, Mike S. Fowler†, Christine A. Johnson‡,

*Department of Mathematics and Statistics, PO Box 68 (Gustaf Hållströmin katu 2b), FIN-00014, University of Helsinki, Finland

†Department of Biological and Environmental Sciences, PO Box 65 (Viikinkaari 1), FIN-00014, University of Helsinki, Finland

‡Evolution, Ecology and Organismal Biology, Ohio State University, 314 Aronoff Laboratory, 318 West 12th Street, Columbus, Ohio 43210, USA

A simple system of checks and balances to cut fraud

SIR — As your Editorial “Standards for papers on cloning” (*Nature* 439, 243; 2006) demonstrates, the fraudulent Hwang stem-cell research papers will have consequences for future research in this and related biomedical fields. As you point out, this does not justify imposing more rigorous standards for reviewing manuscripts in this field than others. Enforcing the deposition of samples with independent laboratories or repositories would be

inappropriate, and deposition could also be done fraudulently. Although deposition might allow another layer of supervision, it would also create another layer of complexity and cost to stymie further research in a field already encumbered by restrictions.

As researchers in different parts of this field, we would like to make a joint response. We believe a simpler system of checks and balances could reduce incidents of scientific fraud and increase our confidence in published reports.

First, all co-authors should indicate the scope of their involvement and declare their understanding of the data in, for example, an author contribution statement such as that recommended by *Nature* (www.nature.com/nature/authors/gta). Surprisingly, it seems clear in retrospect that many of the 26 authors on Hwang's report (*Science* 308, 1777–1783; 2005), could not have attested to the veracity of the human nuclear-transfer embryonic stem cells (ntESC) presented. A requirement for personal accountability might have encouraged greater communication between authors and uncovered the deception before publication.

Second, all journals should, like *Nature* (www.nature.com/nature/authors/policy), require that all published reagents and cell lines be made available to other laboratories.

Finally, peer reviewers should be encouraged to demand that authors provide clear and strong evidence that the data presented support the claims made — including the request for mitochondrial DNA fingerprints if appropriate.

Of course, the best way to ensure integrity in any field is independent replication of results, which requires multiple investigators to be free to do the research. Current limits on US federal funding make independent verification of results especially challenging in the case of human ESC or ntESC research, and undoubtedly contributed to the difficulties in uncovering the misconduct of Hwang and his colleagues.

Xiangzhong Yang

Center for Regenerative Biology, University of Connecticut, 1392 Storrs Road Unit 4243, Storrs, Connecticut 06269-4243, USA

K. Eggen Molecular and Cellular Biology, Harvard University

G. Seidel Jr Animal Reproduction and Biotechnology Laboratory, Colorado State University

R. Jaenisch Department of Biology, Massachusetts Institute of Technology

D. Melton Molecular and Cellular Biology, Harvard University

Peer review could be improved by market forces

SIR — Although peer review seems the best system for quality control of scientific publications and grant proposals (“Three cheers for peers” *Nature* 439, 118; 2006), we might try to improve it. Market forces are

known to optimize complex systems where multiple players have conflicting interests. Economic principles and internet technology could be applied to a peer-review system in the following way. First, a central digital repository receives a paper for a fee ‘x’. Potential referees then bid to review the paper, and, if approved by the author, receive a fee ‘y < x’. Payments are made at the end of each month, allowing for exchanges where an author pays by reviewing other papers.

Referees who can recommend an appropriate journal for the paper and provide the required reference are given due credit and might eventually raise their fee. Authors wanting additional improvements to their work might also pay a higher fee.

Soon, an active exchange could take off where referees quote their position in the peer-market as eagerly as authors quote their citation impact. This system could diminish the workload of referees, by reducing the need to review the same paper for different journals. Eventually this system might be run as an independent peer-review exchange for a profit.

Klaus Jaffe

Laboratorio de Comportamiento, Universidad Simón Bolívar, Apartado 89000, Caracas 1080, Venezuela

Bureaucracy won't change the character of a cheat

SIR — Your Editorial “Standards for papers on cloning” (*Nature* 439, 243; 2006) invites comment on current peer-review procedures and fraud detection. The use of deception for personal gain is neither new nor restricted to human beings. To my mind, the three relevant new things in contemporary society are intense media attention, piles of bureaucracy through which even the lowest-ranking staff have to wade and a culture dominated by lawsuits where redirecting blame seems increasingly prevalent.

During my undergraduate studies, the idea of reporting fictitious data never crossed my mind. And yet I witnessed friends regularly ‘massaging’ graphs, in spite of being taught the proper scientific technique. Is this not, therefore, a problem rooted in personal character? Would the introduction of yet more bureaucracy really solve the problem?

Science and fraud have coexisted for millennia, throughout which time progress was made without computers or armies of administrators. In the case of Hwang and, more recently, Jon Sudbø — who invented test subjects and published his results in *The Lancet* (*Nature* 439, 248–249; 2006) — the open scientific process of peer review, publication and further study revealed the falsehood, and the only people who should be held accountable are those committing the

deception. I believe the scientific method as it exists now is all we need as a community. Indeed, given the current power of the media, the quantity of academic fraud may well decrease, as potential fraudsters witness in full colour the disintegration of their dishonest colleagues.

Phil Bentley

Hahn-Meitner Institut, Glienicke Strasse 100,
14109 Berlin, Germany

Pressure also leads to worthless publications

SIR — Your Editorial “Ethics and fraud” (*Nature* **439**, 117–118; 2006) does not address the problem of ‘publish or perish’.

Researchers are increasingly put under pressure to publish papers to further their career and access resources. But the fact that there are millions of pages published every month, only a few percent of which are worth reading, seems as much a fraud as the Hwang case. Are you wasting your time any more reading something fraudulent than reading something worthless? Neither helps the student or researcher wanting to do something concrete. It seems we have to read ten papers to get the one that really gives us something. The information is fragmented — distributed across hundreds of publications, around the world, many of them inaccessible.

I suggest slowing down the paper-publishing machine by limiting the number of journals that publish original research, asking more peer-reviewers to read preprints and opening up preprint manuscripts for public discussion.

Lindomar B. de Carvalho

International Center for Condensed Matter Physics, Universidade de Brasília, Campus Darcy Ribeiro, Caixa Postal 04513, Brasília, DF - Brazil

It's difficult to publish contradictory findings

SIR — Your recent Editorial (*Nature* **439**, 117–118; 2006) bemoans the recurring subject of ethics and fraud in scientific research. I contend that many journals contribute to the prevalence of bad science, because, when the fundamental observation that led to the original publication cannot be reproduced, it is nearly impossible to publish a paper documenting this. Hence, controversies persist in the literature over many years, simply because the corrected story either is never published, or is not published as prominently as the initial paper.

True, there is an extensive specialist literature where ambiguous or conflicting results can be addressed in detail, but the readership is limited. Some journals, such as *Nature*, have

mechanisms for publishing technical comments on published research (Brief Communications Arising: online only; see www.nature.com/nature/authors/gta/briefcomms.html#a2), but these are few in number and must adhere to strict criteria.

Reviewers of contradictory results often ask that the authors explain how the original authors could have obtained their results. To quote a recent rejection letter, “an adequate explanation for the apparent contradictory findings is not provided”. Certainly, speculative explanations can be offered for some kinds of experimental differences. But it is never possible to prove how another lab obtained data that cannot be reproduced. One can only be certain of one's own data. This demand for explanation creates serious problems in the case of scientific fraud. In a minor case, the original authors may have fudged one small set of data to ‘prove’ their theory. In a more serious case, fundamental observations cannot be reproduced. Whether this irreproducibility is due to outright fraud, scientific incompetence or some combination cannot be determined by the authors who try to reproduce the result and fail.

Another often-made request of reviewers is that the original experiments be reproduced exactly. This sounds reasonable but, in fact, can become an absurd burden. Even if the methods section were complete and accurate, one can never say with certainty that one has reproduced the experimental conditions precisely. Instead, the appropriate approach is to design experiments to test the conclusions of the original paper. If these conclusions are disproved, then the details of how they were arrived at are not relevant.

Of course, a contradictory paper should be held to a higher standard than was the paper it refutes. But all journals must endeavour to correct errors, or those who perpetrate scientific misconduct (not necessarily outright fraud) will be rewarded, and those who try to correct wrong hypotheses in the proper hegelian manner — thesis, antithesis, synthesis — will be punished.

Thomas E. DeCoursey

Molecular Biophysics and Physiology, Rush University Medical Center, 1750 West Harrison Street, Chicago, Illinois 60612-3824, USA

Data audit would reduce unethical behaviour

SIR — I agree with the point made in your Special Report “Should journals police scientific fraud?” (*Nature* **439**, 520–521; 2006) that editorial offices are not the proper place to monitor fraud. This is why, 19 years ago, my colleague Zoltan Aannau and I proposed data audit (*Nature* **327**, 550; 1987).

Research subject to data audit could include studies presenting possible risks to public

health, or those questioned by a whistleblower or by peer review. Others could be subject to random audits. Up to 1% of all studies could be audited every three to five years, at less than 1% of the cost of the original study (*Accountability Res.* **1**, 77–83, 1989). Auditing could be done by an independent body that would certify the validity of published results. Sponsoring institutions could choose to publish a transparent analysis of selected papers on the web.

Although these processes might not eliminate all fraud or misconduct, they could substantially reduce such unethical practices.

Adil E. Shamoo

Accountability in Research, University of Maryland School of Medicine, 108 North Greene Street, Baltimore, Maryland 21201, USA

Discourse among referees and editors would help

SIR — In the discussion on enhancing peer review, following publication of the controversial human-cloning papers in *Science* (“Ethics and fraud” *Nature* **439**, 117–118; 2006), I would like to highlight one of the limitations of this process. Last year, I was asked to review a manuscript for a high-impact journal. Although I duly reviewed the paper — before the deadline and after extensive reading and research — I have yet to receive information on the paper's status.

I see three possible scenarios regarding the paper's fate: either it has been accepted or it has been rejected or the authors have been advised to revise it. In the first and second cases, a referee likes to know how an editor made their decision in light of, or in spite of, any objections raised. In the third case, each referee likes to know what comments or recommendations other referees and the editor have made, as well as details of the authors' rebuttal. Obviously each referee is an expert in his or her field, but not necessarily in the other sub-fields relevant to a particular manuscript. Sharing the referees' comments is essential to the learning (and in some cases validation/checking) process. It might also help clarify differences of opinion between referees of the same expertise. I believe discourse among the referees and editor would enrich the quality of reviewing and might prevent scandals such as the human stem-cell line cloning debacle.

Debomoy K. Lahiri

Institute of Psychiatric Research, Indiana University School of Medicine, 791 Union Drive, Indianapolis, Indiana 46202, USA

Nature editors aim to inform all peer reviewers when a final decision about publication of a manuscript is made, and to send reviewers each others' reports on the manuscript, together with a letter of thanks — Editor, *Nature*.

COMMENTARY

Mapping disaster zones

Google Earth software proved effective during relief efforts in New Orleans and Pakistan, say **Illah Nourbakhsh** and colleagues. Is there more to be gained than lost from opening up disaster operations to the wider public?

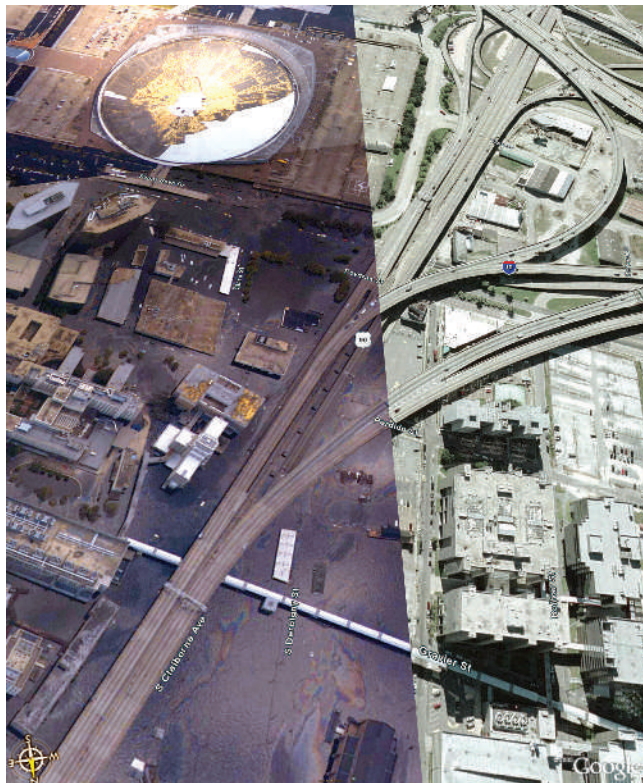
In the aftermath of any disaster, satellite and aerial images are critical for identifying priorities, planning logistics and working out access routes for relief operations. Two events last year, Hurricane Katrina in the United States and the Pakistan earthquake, catalysed a collaboration between relief workers and the Global Connection project, a partnership between Carnegie Mellon University in Pittsburgh, Pennsylvania, NASA/Ames Research Center in Moffett Field, California, Google and National Geographic. The goal was to help communities and response agencies to access post-disaster images quickly using the Google Earth geospatial image browser. The emergence of a new breed of volunteers — online data managers — highlights the potential of a web-based community approach to disaster operations.

These experiences have taught us that real-time spatial image browsing can have huge benefits to aid operations. But the technology was used in a markedly different manner after Hurricane Katrina from the way it was used after the Pakistan earthquake, owing to greater technical and economic barriers in the developing world. The use of government data sources can be restricted for security reasons, and although this is true everywhere, public pressure for transparency is usually higher in the developed world.

Earth from above

Satellite images from commercial sources are not free, and are available to the public only when a third party, in this case Google, chooses to purchase and release the data. Software tools for professional image analysis and processing are also expensive. Future developments that make these tools available to a wider audience may lead to sacrifices in privacy and security. We believe such sacrifices are inevitable and can be justified by both direct and indirect humanitarian benefits.

The Google Earth model of the globe consists of hundreds of thousands of individual satellite and aerial images taken from more



Aerial photograph of New Orleans before and after Hurricane Katrina.

than one hundred data sources, including NASA's Landsat satellite, commercial satellites such as Digital Globe's QuickBird, and providers of aerial photographs. These images are stitched together to create a spatially searchable, high-resolution view of Earth. In urban areas, resolution may be as high as 6 to 12 inches. In other regions, particularly rural areas and developing countries, only 15-metre Landsat resolution may be publicly available.

This low resolution is one of the obstacles facing spatial image browsing for disaster response. Many of the tasks that relief teams encounter, such as identifying roads with crack-free surfaces after an earthquake, require images with one-metre resolution or better. A second obstacle is image freshness. Images seen while browsing Google Earth can be between six months and three years old, depending on their source. However, by making use of updated overlays provided in a format known as Keyhole Markup Language

(KML), it is possible for multiple users to update Google Earth in real time. In this way, online volunteers can help to overcome resolution and image-freshness issues.

A patchwork picture

Global Connection's original mission with National Geographic produced 500 high-resolution image overlays of Africa derived from aircraft photographs taken in 2004. When viewing Africa using Google Earth, small red aeroplane icons appear; zooming in for a closer look reveals a high-resolution aerial image overlay on top of the Landsat background image.

This technique of overlaying images found an unexpected application when Hurricane Katrina flooded New Orleans and the surrounding areas on 29 August 2005. The National Oceanic and Atmospheric Administration (NOAA) captured more than 8,000 images of flood-damaged areas over 10 days using a high-resolution camera mounted on a small high-speed aircraft. This information was invaluable

to rescue efforts, and in the first week some 5 million photos were downloaded from NOAA's website each day.

But navigating this huge, unprocessed data set in a meaningful way required the images to be presented in a searchable, stitched-together format. At the request of NOAA, Google Earth and the Global Connection team created new KML software tools to handle the images that NOAA released on a daily basis. These tools made the preparation of image overlays faster than was previously possible. During September, we estimated that the Global Connection server supported some 2.5 million viewings of the high-resolution overlays of post-Katrina images by users of Google Earth.

In the case of Hurricane Katrina, spatial image browsing proved useful to federal disaster agencies and, we believe, to the larger public. The technology personalized disaster data by making each individual an active explorer, capable of discovering information that was relevant to them. For instance, Google Earth

IMAGE FROM GOOGLE EARTH/SANBORN; OVERLAY BY NOAA & GLOBAL CONNECTION PROJECT

images of flood-free suburbs in New Orleans helped some families in Pennsylvania to identify intact churches that could be used as an outlet for food donations.

As the Global Connection team converted the last of the NOAA images for Hurricane Rita — which reflooded New Orleans on 23 September — to overlays in October 2005, volunteers in Pakistan were already responding to the 8 October earthquake and collecting disparate satellite data of the disaster zone. Specialized mapping outfits, such as the UK charity MapAction (www.mapaction.org), began working with the United Nations in Islamabad to distribute satellite images and information maps to agencies on the ground. Even though most publicly available images had coarse resolution (limited to 15 metres), MapAction was able to collate essential logistical information, such as major road blockages and the locations and numbers of tents required throughout the area.

Some members of the disaster response community in Pakistan who had heard of Google Earth contacted Global Connection in the hope of finding higher-resolution images. An e-mail on 12 October from Ayaz Abdulla of The Citizens Foundation (TCF), a Karachi-based organization that switched their mission from education to disaster relief when the earthquake happened, explains their need for higher-resolution data:

"I have the basic version of Google Earth which shows the terrain to some detail, but our relief workers (and logistical strategists) would be monumentally helped by images with higher resolution (especially with better clarity of roads, and so on). The TCF relief effort currently continues with limited and outdated maps that don't truly reflect the current terrain (and more importantly, the inhabited areas). Local know-how of the region is extremely useful, but I feel that this satellite technology could be priceless."

Beyond the sky

Unlike the situation after Katrina, there was no high-speed aircraft surveying the disaster area in Pakistan nor high-resolution images available, apart from commercial satellite data. But by 14 October, Google Earth had acquired, processed and published two satellite images from Digital Globe's QuickBird that were taken on 9 October, the day after the earthquake. In the five days that followed, high-resolution images taken by the commercial IKONOS satellite were released publicly on several websites, then withdrawn following Pakistan's concerns over security in Kashmir, and finally re-published by the United Nations following talks with Pakistan and India (www.nature.com/news/2005/051017/full/051017-12.html).

Even when more images became available,

most were not immediately usable due to lack of computer-readable information about the projection and location of the image, known as a 'world file'. The NOAA images of Katrina recorded the aircraft's position and attitude, allowing automatic generation of world files. Fortunately, the best IKONOS images available online, prepared by the Center for Satellite-Based Crisis Information in Munich,

Germany, were published with world files. Global Connection used these to generate one-metre resolution image overlays of parts of the districts of Muzaffarabad, Mansehra and Abbottabad by 24 October.

The IKONOS images yielded updated overlays for Google Earth for only a limited region, but the subsequent feedback from relief workers in the field, in this case the TCF, was extremely useful to the US-based Global Connection team. The TCF helped us to prioritize what areas of Pakistan to focus on next. Given the project's limited resources and the sudden flood of public data now available, this was essential for focusing the team's efforts.

We learned many other lessons from Pakistan, not least the need to develop tools that can adapt to local conditions on the ground. Internet connections in Pakistan were often slow and patchy, making downloads difficult. Post-disaster feedback highlighted difficulties with printing out maps and locating specific settlements among tens of thousands. Another obstacle was the mismatch between the local situation and the way existing geographic data is organized and represented.

For example, Pakistani villages do not have a single position, but exist along a right-of-way extending from the valley to a ridge top, migrating up and down based on the changing seasons. Thus assigning a fixed position to a village is wrong most of the time. If Google Earth could label a village boundary, rather than a single location, this might prevent settlements being misidentified.

More generally, the connection between city or village name and position can be flawed. Spatial image browsers such as Google Earth make use of several databases that link place names with latitude and longitude, and unfortunately all such databases suffer from inaccuracies, such as placing Islamabad 100 miles from its true location. All mapping efforts, whether community based or UN supported, would benefit from a stronger feedback mechanism, such as a community filtering system that enables authenticated individuals to update shared geographical data online. For example, we hope that database providers will support initiatives, such as the Geonames Forum (www.geonames.org/about.html).

Where will spatial image browsing take

disaster response in the future? We hope that a growing awareness of such browsers will encourage greater standardization of image and data formats, including world files. This would work best with the support of a reputable organization, such as the US National Institute for Standards and Technology. Low-cost global positioning systems combined with automated image analysis could allow browsers such as Google Earth to create overlays from images taken with cheaper instruments; for example, cameras in mobile phones or camcorders in helicopters. In time, we hope to see the public's role shift from passive viewer to active contributor.

Global contribution

This prospect will no doubt raise privacy concerns. However, immediately after a disaster, few would deny the value of making high-resolution, up-to-date imagery public for rescue operations. But we would argue that the benefits go beyond the immediate logistics of disaster response. Individuals, moved by pictures, may make donations and encourage their governments to do the same. There are many organizations and individuals who can contribute in unexpected ways following a disaster. But there is also a danger of overwhelming relief operations with tens of thousands of enthusiastic image donors.

There are real challenges in scaling up such activities: from validating the accuracy of contributions to standardizing image formats. An online organization, similar to MapAction in its goals, is needed to serve as a bridge between the sourcing of high-resolution satellite data

and humanitarian agencies on the ground. Moreover, we believe that the highest-resolution data, such as that from Digital Globe, should be rapidly made public following a disaster. For instance, the International Charter on Space and Major Disasters (www.disasterscharter.org) is a collabora-

tion that provides digital maps sourced from space-agency satellites. We need to strengthen organizations such as Europe's Respond consortium (www.respond-int.org/Respond) that can acquire commercial high-resolution images, and encourage standardized source data. Such advances, combined with image-overlay technology, promise to make spatial image browsing far more timely and effective in relief efforts. ■

Illah Nourbakhsh and Randy Sargent are at Carnegie Mellon University, Pittsburgh, Pennsylvania; Anne Wright is at NASA/Ames, Moffett Field, California; Kathryn Cramer; Brian McClendon and Michael Jones are at Google Earth. For more on Google Earth see pages 763 and 776.

Acknowledgements Many at Google, Carnegie Mellon, NOAA and NASA are acknowledged for this effort; see www.cs.cmu.edu/~globalconn/rricredits.html for recognition of these individuals.

BOOKS & ARTS

Enough is enough

Economic growth should be viewed as a means to an end, not an end in itself.

The Logic of Sufficiency

by Thomas Princen

MIT Press: 2005. 401 pp. \$29

Robert Costanza

How much is enough? This is a question that is almost never asked in conventional discussions about the economy and society. The tacit assumptions are that there is never enough, that growth must continue indefinitely, and that, as Robert Heilbroner once put it: “the flow of goods and services consumed by everyone constitutes the ultimate aim and end of economic life”. These assumptions are now so ingrained that to question them seems blasphemous, but a little research shows that they are fairly recent (post-Second World War) and are ultimately counterproductive in a finite (and full) world. Thomas Princen, associate professor of international natural resources and environmental policy at the University of Michigan, has added his voice to the growing chorus recognizing the ultimate futility of unlimited economic growth on a finite planet.

Princen's unique contribution to this discussion is his detailed and engaging history of the efficiency principle and its role in supporting the paradigm of unlimited economic growth. With its roots in factory management, efficiency has grown to encompass all aspects of modern life, Princen says. But efficiency requires specialization and a large scale, and for many workers this results in the removal of the elements of a job that make it fulfilling. Those elements include creativity, control and responsibility. Princen contrasts this with a proposed sufficiency principle, which places ‘good’ above ‘ever better’ and aims to provide a rationale for quality and sustainability.

He provides three detailed contemporary examples where the logic of sufficiency has been implemented and the systems have managed to survive and do well, while still embedded in the larger efficiency- and growth-crazed society. The first is the northern California Pacific Lumber Company's sustainable timber-harvesting policies; the second is the protected and restricted lobster fishery at Monhegan Island, Maine; the third is Toronto Island's policy of restricted car access. In each of these chapter-length examples, Princen details what can be achieved in terms of quality of life and sustainability by breaking with the idea that more is always better. The examples

also provide an existence proof for the sufficiency principle — solutions do exist.

Princen argues that if we adopt a “logic of sufficiency” rather than efficiency, the larger culture can also achieve these goals. There is much to like in this book, but it sets up a false dichotomy between efficiency and sufficiency, and ultimately misses the core issue — it is alternative goals, not just alternative rationalities, that lead to different results. The logic of efficiency can just as easily be used in the service of sustainability as in the service of growth.

The core problem is that we have forgotten that economic growth is a means to an end, not an end in itself. What, then, is the end? What is the shared goal of humanity? There is a growing consensus that this goal is (or should be) a sustainable and desirable present and future for all of humanity. It is not enough for the system to be merely sustainable, as an atrocious system might be sustainable indefinitely. Likewise, it is not enough for the system to be merely desirable, as desirability now may lead to misery later. It is also not enough for the system to be desirable in some of its aspects, but terrible in others. For example, economic development that causes environmental and social destruction that outweighs its gains is not real development at all. Finally, it is not enough for the system to be sustainable and desirable for a small élite, while leaving most people in sustainable misery.

If our goal is sustainable human well-being,

we have to measure and include all the things that contribute to that goal. These not only include conventional economic goods and services, but also contributions from nature, from family, friends and other social relationships at many scales, and from health, education and fulfilling employment. One convenient way to summarize these contributions is to group them into four basic types of capital that are necessary to support the real, human-welfare-producing system: built capital, human capital, social capital and natural capital. Given the increasingly well known dependence of human well-being on ecosystem services and natural capital, this implies that maintaining our ecological life-support system and the biodiversity that allows it to function is a key sub-goal. Nature is not merely a luxury good.

Ultimately, to answer the question ‘How much is enough?’, one needs to specify ‘for what purpose — towards what goal?’. If the purpose is ‘more’ then, obviously, there can never be enough. But if the goal is a sustainable and desirable future for all of humanity, then ‘enough’ has a clear meaning and imperative. Our challenge is to clearly articulate and agree on the vision of a sustainable and desirable future for humanity. Then we can clearly define what is sufficient to meet that goal, but we can also be efficient in achieving it. ■

Robert Costanza is at the Gund Institute of Ecological Economics, University of Vermont, Burlington, Vermont 05405-1708, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Protecting resources such as Borneo's rainforests can help give humanity a sustainable, desirable future.

Super science

The Physics of Superheroes

by James Kakalios

Duckworth: 2006. 384 pp. £9.99

Death Rays, Jet Packs, Stunts and Supercars: The Fantastic Physics of Film's Most Celebrated Secret Agent

by Barry Parker

Johns Hopkins University Press: 2005. 231 pp. \$25

A Teaspoon and an Open Mind: The Science of Doctor Who

by Michael White

Allen Lane: 2005. 192 pp. £12.99

The Science of Doctor Who

by Paul Parsons

Icon Books: 2006. 240 pp. £12.99

Stephen Baxter

Comic-book superheroes might seem an unlikely starting point for a popular book on physics. As James Kakalios relates, the ur-hero Superman was born in 1938 as a Depression-era revenge fantasy — the Man of Steel's first enemies were corrupt landlords and Washington lobbyists. And in the tough darwinian world of pulp comics, the aim of story-telling is to make you turn the page, not scientific accuracy.

But as Kakalios points out, the exploits of a superhero can illustrate scientific principles, although you may need to make a 'miracle exception'. Once you accept that Superman's stablemate the Flash can somehow run at arbitrarily high speeds, then you can study the consequences, such as traction, deceleration forces and nutritional requirements. Indeed, sometimes the comic-book writers explore the science themselves. Kakalios quotes an adventure in which the Flash, in order to save the citizens of a North Korean city from a nuclear blast, runs at close to the speed of light and suffers relativistic effects: "As his body sloughs off the screaming after-effects of near light travel, eyes of almost infinite mass turn towards the blaze engulfing Chongjin." There can be poetry in the physics.

The 'science-of' the latest popular franchise has been a flourishing subset of the popular-science genre since the success of Lawrence Krauss's *The Physics of Star Trek* (Basic Books, 1995). You can see the appeal for authors and publishers. Fans can be seduced through their curiosity about the infrastructure of their favourite universe — could a machine really travel through time? could a man really fly? — into explorations of genuine science. There is thus a benignly educational motive. And, of course, you can sell an awful lot of books to all those fans.

But to get it right you have to focus on the needs of the readership: a fan wants to read a book about the franchise, not a textbook. Of the new crop of such books reviewed here,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Kakalios's work on superheroes is probably the most successful. He admits to being a fan himself, the text is drenched with fan-friendly references to the comics, and the physics, even when Kakalios points out where the comics got it wrong, is drawn out sympathetically and with good humour.

Barry Parker's study of James Bond's stunts and gadgets is thorough and clearly written, but is something of a plod. You can read about the physics of car chases, but if you're hoping for wild speculation on Bond's invisible car (in the movie *Die Another Day*) you'll be disappointed. Studded with a lecturer's stories from the classroom, it also has a whiff of chalk dust.

Such is the current popularity of the television series *Doctor Who* in Britain that it's no surprise to see two new 'science-of' books on the subject. Michael White's tome has an elegant title, based on a Tom Baker line from the show itself: "Well, to be fair, I did have a couple

of gadgets he probably didn't, like a teaspoon and an open mind." But there simply isn't enough about the Doctor: White's elegantly written but run-of-the-mill essays on time travel and alien life contain only glancing references at top and tail to the Time Lord. Paul Parsons is the editor of the BBC's *Focus* magazine, and his contribution is snappy, lively, journalistic, has sound bites by various tame boffins, and is drenched in Doctor Who. It is more imaginative too, with explorations of off-beat topics such as the science behind the Doctor's two hearts and his altruism.

The most interesting 'science-of' series is probably that accompanying Terry Pratchett's Discworld books. Written by Jack Cohen and Ian Stewart with Pratchett, the latest is *The Science of Discworld III: Darwin's Watch* (Ebury Press, 2005), an exploration of darwinism. These books are unique (to my knowledge) in that they contain contributions by the author of the franchise itself. And the authors' intention isn't just to deliver more pop-science books, but to develop ongoing scientific arguments (notably Cohen and Stewart's hypotheses on complexity and intelligence) in a popular form. In that sense these books reach back to an age in which scientists were expected to express their arguments in a form comprehensible to the layman; this is science being done in public.

Not all franchises lend themselves to scientific explorations — not that this deters the attempt. Anyone tempted by *The Science of Harry Potter* (Roger Highfield; Viking, 2002)? But as a former secondary-school teacher, I can testify to the usefulness of pop-culture examples to snag the student imagination. Spider-Man swinging on his web is a compelling example of a simple pendulum.

Even if they don't snare the fans directly, 'science-of' books can serve as a useful resource for teachers, and when well done can make a valuable contribution to the public understanding and awareness of science. ■

Stephen Baxter's latest science-fiction novel is *Transcendent* (Gollancz, 2005).

Incomplete mathematics

Meta Math! The Quest for Omega

by Gregory Chaitin

Pantheon Books: 2005. 223 pp. \$26

Ivor Grattan-Guinness

The surprise of Kurt Gödel's incompleteness theorem of 1931 lay not so much in the incompleteness itself, but that it was found in so simple a mathematical theory as first-order arithmetic. It follows, then, that any richer mathematical or logical theory, which is most of them, is also incomplete. In addition, Gödel's method of proof, elaborating what became known as recursive functions, was fruitful in

its own right. In particular, it helped in Alan Turing's creation in 1936 of computable numbers and his finding that it cannot be decided in a finite number of steps whether or not a computer can calculate some given number, and whether or not any formula expressible in this system is also a theorem of it.

Turing is Gregory Chaitin's hero, to the extent of being credited with the more general result that it cannot be decided whether or not a computer will complete a given task in some finite number of steps. This extension of Turing's conception, which creates the 'halting problem', actually seems to be attributable to

EXHIBITION

Classifying the past

The photograph of artist Mark Dion included in his latest exhibition, *Microcosmographia*, is strikingly similar to one of the American naturalist William Beebe taken in 1917. This is no accident: Dion's adoption of the attitudes and methods of such early naturalists is very deliberate. Through *Microcosmographia*, Dion highlights problems of accuracy, past and present, in natural history and taxonomy.

The exhibition's central piece, *Ichthyosaur*, pictured here, references the confusion of palaeontological classification in the early nineteenth century. Various misinterpreted by fossil collectors and palaeontologists as prehistoric fish, predecessors of modern crocodiles, or relatives of the duck-billed platypus, the genus *Ichthyosaur* was only officially designated as such by William Daniel Conybeare in 1822.

The belly of Dion's life-size resin replica of a beached ichthyosaur is split open and overflows with

the paraphernalia of early naturalists, ranging from old reference manuals to glass beakers. The work seems to suggest that the ichthyosaur is a creature quite literally made up of the past; that its inner workings are defined by the humans who discovered it and eventually classified it. The concept resonates with present-day taxonomy, which is struggling to systematize 250 years of natural history of varying scientific quality: sifting out errors, identifying missed connections, and establishing a comprehensive informatics for the field.

Confusions and corrections, such as those surrounding the ichthyosaur, inspire Dion's work. The diverse collection of sculptures, drawings and photographs in this exhibition, and their juxtapositions and groupings, focus on long-dead scientists and their influences on scientific understanding today. But the collection also raises important



questions about the fallibility of science. Dion seems to want to point out the mistakes of the past in order to warn us about mistakes we might be making now, and might make in the future.

Microcosmographia, organized by the South London Gallery, can be seen at The Harris Museum and Art Gallery in Preston, UK, until 12 March 2006.
Alexis Clements

SOUTH LONDON GALLERY

Martin Davis in the early 1950s. In *Meta Math*, Chaitin has extended this line of thought by using the notion of the computer program to study incomplete mathematical theories. Written for the general reader, the book consists of a main text of about 160 pages, followed by reprints of two earlier papers of a more technical character.

A key notion in this book, inspired by biology, is the 'complexity' of a program, specified in terms of the smallest program(s) (in bit size) that can produce some given output. Such a program is irreducible, or incompressible: "So you can define randomness as something that cannot be compressed at all," according to Chaitin. Thence follows the Omega number, an infinitely complex positive real number specifying the halting probability. As Chaitin puts it, it is defined from all the programs chosen by chance to run on a fixed computer and also to continue to run by random operator decisions until the computer "must decide by itself when to stop reading the program"; if a program halts after k bits, then it contributes $1/2^k$ to the number. As stated, the number depends on the computer used; presumably this machine is Turing-powerful enough to run any program installed on it. Drawing upon a discussable claim that any mathematical theory can be encoded in programming terms, Chaitin

concludes that the Omega number "marks the current boundary of what mathematics can achieve".

Chaitin's investigation is attended by several skirts around paradoxes, especially those involving naming. For example, to qualify as a program contributing to the Omega number, the program has to be able to say how large it is when it halts. This type of concern also owes much to Gödel's 1931 theorem, where new standards were imposed on distinguishing logic from metalogic. It is a pity that Chaitin never states that theorem precisely, and once even states it quite wrongly.

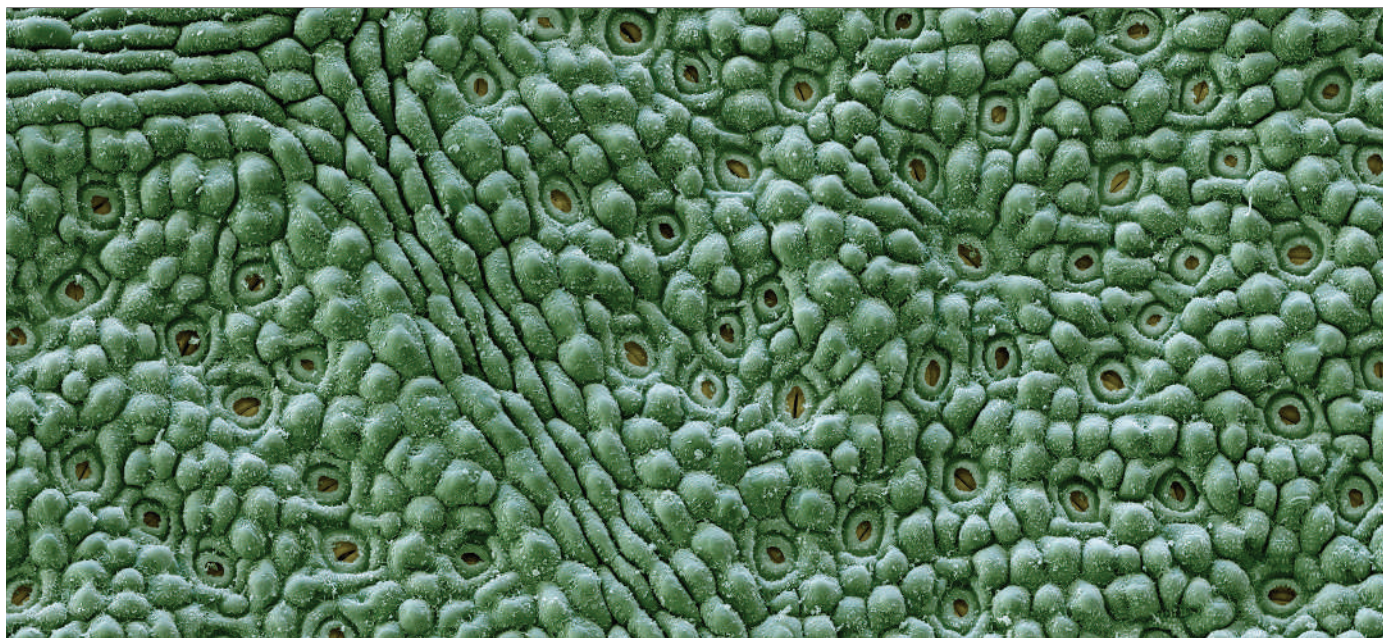
The scope of the author's meta-programme (as it were) is impressive: essentially straightforward assumptions and steps lead to some wide-ranging consequences and claims about mathematics, logic and computing science. The account is nicely signposted by the frequent use of information boxes containing the main definitions, steps or relationships. As the book is intended for a wide audience, it might have been enriched by some comments on concurrent developments that have used versions of the main notions; for example, (non-biological) complexity with A. N. Kolmogorov in the 1960s, or the realm of intelligent activity lying beyond computability, as debated by Roger Penrose and others in recent times.

The style of writing throughout is better suited to an internet chatroom than to a book ("Discours de Métaphysique — that's the original French" is only one such example) and has exclamation marks spread liberally. Instead of properly referencing works that are precisely cited in the text, "I decided to concentrate mostly on recent books that caught my eye," says the author. The list lacks, among other key works, J. W. Dawson's *Logical Dilemmas*, *The Life and Work of Kurt Gödel* (A. K. Peters, 1997) and *The Essential Turing*, edited by B. J. Copeland (Oxford University Press, 2004).

Many historical remarks are made, but are seemingly free of knowledge of the figures involved and their importance. For example, "the nearly-forgotten 17th-century genius Leibniz", "Newton's incomprehensible *Principia* — written in the style of Euclid's *Elements*", or "it was Cantor's obsession with God's infiniteness and transcendence that led him to create his...theory of infinite sets and infinite numbers". The reader should be ready to add their own exclamation marks to such passages.

It is nice to have popular books on modern mathematics, logic and science. But it is nicer if they are prepared with care.
Ivor Grattan-Guinness is at Middlesex University, Enfield, Middlesex EN3 4SF, UK.

NEWS & VIEWS



EYE OF SCIENCE/SPL

GLOBAL CHANGE

The water cycle freshens up

Damon Matthews

Rivers are delivering increasing amounts of fresh water to the ocean. The cause seems to be the influence that higher concentrations of atmospheric carbon dioxide are having on water use by plants.

Measurements of stream flow around the world have documented an increase in the amount of water that runs off the continents and returns to the ocean¹. This trend has been occurring since the beginning of the century, yet changes in precipitation over land do not sufficiently account for this increase. On page 835 of this issue, Gedney *et al.*² identify an important contributor to increasing global runoff — decreased evaporation resulting from the influence of elevated atmospheric carbon dioxide on plant physiology (Fig. 1, overleaf).

Carbon dioxide is the currency of plant photosynthesis: plants take up CO₂ from the atmosphere and incorporate it into their tissues in the form of organic carbon compounds. This uptake of CO₂ is accomplished through plant stomata — small openings in the surface of leaves (pictured above) that open and close to allow the exchange of CO₂ and other gases with the atmosphere. During gas exchange, water is inevitably lost to the atmosphere, again through stomatal openings. This is the process of plant transpiration, which, on a global scale, mediates the transfer of water from the soil into plant tissues, and out through stomata to the atmosphere. On vegetated land, plant

transpiration can make a substantial contribution to total surface evapotranspiration, which represents the sum of plant transpiration and other surface evaporation.

Plants can regulate the opening and closing of stomata in response to changing environmental conditions; in a high-CO₂ atmosphere they are more efficient in their use of soil moisture. The stomata do not open as much or for as long, and less water is lost from leaves to the atmosphere³. As a consequence, plants acquire enough carbon through their stomata with less water uptake from the soil. The result is that continental evapotranspiration is reduced, more moisture is left in the soil, and this additional surface water can lead to increased continental runoff⁴.

Using a technique known as 'optimal fingerprinting' (also known as 'detection and attribution'), Gedney *et al.*² show that this direct effect of elevated CO₂ on plant transpiration is the dominant contributor to observed increases in continental runoff. Optimal fingerprinting is simply a statistical regression in which a model simulation is compared with observations to isolate which processes in the model are consistent with the observed data. If

a model simulation is consistent with observations, the process that drives the model trend is said to be 'detected' in the observations; if the observed trend is also inconsistent with other plausible explanations, then the trend can be 'attributed' to a specific cause.

Gedney *et al.* investigated four plausible contributors to observed increases in runoff: climate change leading to changes in temperature and precipitation; land-use change and consequent changes in vegetation cover; so-called 'solar dimming', resulting from an increasingly hazy atmosphere; and the direct effect of CO₂ on plant transpiration. The effects of each of these on surface runoff were simulated using a sophisticated land-surface and vegetation model, and the results of the model were compared with historical observations of continental runoff. The authors' analysis shows that model-simulated runoff trends are consistent with the observed trend only when the direct effect of CO₂ on transpiration is included in the simulation. So they attribute increases in continental runoff over the past century to the physiological effect of elevated atmospheric CO₂.

Detection and attribution has been widely

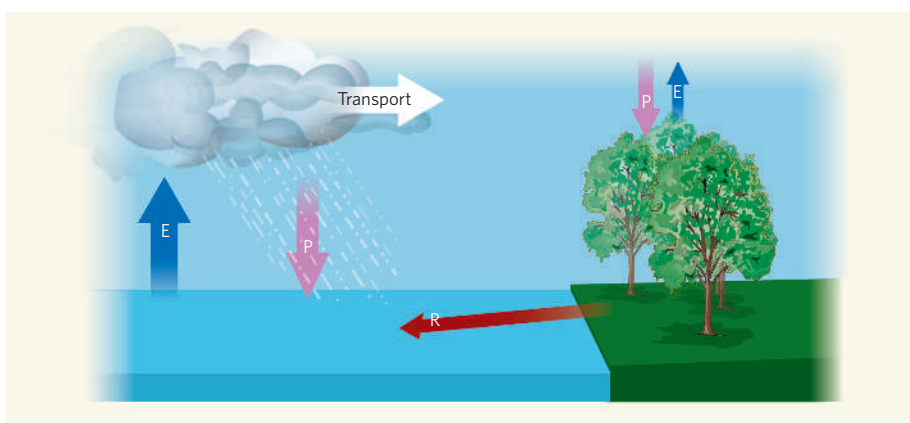


Figure 1 | Plants, CO₂ and the global water cycle. The balance between precipitation (P) and evaporation (E) over land determines the surface runoff (R), which returns water from the continents to the oceans. Plant photosynthesis plays an integral role in the global water cycle, by mediating the transfer of water from the land surface to the atmosphere. Elevated CO₂ can lead to closure of leaf stomata, which reduces leaf water loss and thereby decreases overall continental evaporation. Gedney *et al.*² show that this process, initiated by increased atmospheric CO₂, can account for the increases in surface runoff observed over the past century.

used in climate science to attribute observed climate trends to both natural and anthropogenic causes⁵. Recent increases in surface temperature have been successfully attributed to increases in CO₂ and other greenhouse gases⁶, as have changes in other climate variables such as ocean heat content⁷ and sea-level pressure⁸. Gedney and colleagues' research represents the first time that detection-and-attribution techniques have been used to

successfully link changes in the functioning of terrestrial ecosystems to human influences on the atmosphere.

As with any statistical analysis, these results are only as sound as the model used, the experimental design and the quality of observations. As our understanding of the terrestrial biosphere and our ability to model and monitor it improves, contributors to observed runoff trends that were not considered in this

study may well be identified. Gedney and colleagues' findings are nonetheless an important step forward in our understanding of the diverse and complex ways in which human activities are affecting the global climate system. The findings have implications for future surface warming and freshwater availability, both of which could increase if CO₂ continues to affect surface-water fluxes as demonstrated here. This research also opens some fascinating avenues of investigation — such as the possibility of using records of river runoff to monitor the functioning of terrestrial ecosystems in response to climate change, or of studying how changes in runoff induced by elevated CO₂ might affect ocean circulation.

Damon Matthews is in the Department of Geography, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada.
e-mail: dmatthew@ucalgary.ca

1. Labat, D., Godderis, Y., Probst, J. L. & Guyot, J. L. *Adv. Wat. Res.* **27**, 631–642 (2004).
2. Gedney, N. *et al.* *Nature* **439**, 835–838 (2006).
3. Medlyn, B. E. *et al.* *New Phytol.* **149**, 247–264 (2001).
4. Field, C., Jackson, R. & Mooney, H. *Plant Cell Environ.* **18**, 1214–1255 (1995).
5. International Ad Hoc Detection and Attribution Group *J. Clim.* **18**, 1291–1314 (2005).
6. Tett, S. F. B. *et al.* *J. Geophys. Res.* **107**, 4306 (2002).
7. Reichert, K. B., Schnur, R. & Bengtsson, L. *Geophys. Res. Lett.* **29**, 1525 (2002).
8. Gillett, N., Zwiers, F. W., Weaver, A. J. & Stott, P. A. *Nature* **422**, 292–294 (2003).

DEVELOPMENTAL BIOLOGY

Two paths to silence merge

Panthea Taghavi and Maarten van Lohuizen

To maintain their identity across generations, specialized cells must heritably repress swathes of genes — keeping active only genes necessary for the cell's purpose. Now it seems two repressive pathways join forces.

Even though cells of all developmental stages carry the same DNA, they have their own identity, defined by the combination of proteins expressed in each cell. These expression patterns, although set early during development, are reproduced in each mature, specialized cell later in life, over many cell divisions. This phenomenon is referred to as 'cellular memory'. Any perturbation of this fine-tuned system is a hazard to proper development and health.

Cellular memory is thought to be regulated by two 'epigenetic' mechanisms, which heritably change the characteristics of the cell without altering its DNA sequence. These mechanisms do this by altering the chromatin (the DNA and its associated proteins), which changes the availability of the genes to be expressed. One mechanism involves a group of enzymes that attach methyl groups to the DNA, the DNA methyltransferases (DNMTs);

the other involves Polycomb group (PcG) proteins, which modify the histone proteins around which the DNA is wrapped. On page 871 of this issue, Viré *et al.*¹ show a direct interaction between DNMTs and the Polycomb protein EZH2 that hints at how these complex systems might collaborate to set up cellular memory.

DNA methylation helps to reorganize chromatin into a 'silent' state in which genes are not expressed. In mammals, DNMT3A and -3B are mainly responsible for establishing methylation at previously unmethylated sites, whereas DNMT1 is the major maintenance methyltransferase, reproducing existing methylation patterns during cell division².

PcG proteins were originally identified for their role in maintaining the repressed state of certain developmental genes in the fruitfly *Drosophila*. EZH2 is a histone methyltransferase

that can methylate lysine 27 of the tail of histone H3 (H3-K27) or lysine 26 of histone H1 (H1-K26)³. This enzyme functions as part of the Polycomb repressive complexes 2 and 3 (PRC2/3), and whether it methylates H1 or H3 depends on which other proteins are in the complexes⁴. These histone modifications can attract other repressive complexes, such as Polycomb repressive complex 1 (PRC1), that then propagate the silenced state^{4,5}.

In their study, Viré *et al.* show convincingly that EZH2 interacts with all three DNMTs *in vitro* and *in vivo* (Fig. 1). DNMTs have not been found previously in purified PcG complexes from *Drosophila* or mammals, perhaps indicating that the interactions observed by Viré *et al.* are transient. How could these interactions affect the gene-expression programme of the cell? The authors find that proper repression of a few genes targeted by EZH2 requires both EZH2 and DNMTs. Moreover, EZH2 is needed to bring DNMTs to the regulatory regions ('promoters') of the EZH2 target-genes. Depletion of EZH2 disturbs this recruitment, derepressing the genes and allowing expression. At the same time, there is a decrease in the repressive histone H3-K27 methylation marks.

These results reveal that EZH2, as part of the PRC2/3 complexes, can physically recruit DNMTs to certain target-genes and that this process is essential for silencing the genes.

Surprisingly, reducing the levels of any of the DNMTs separately results in similar de-repression of the EZH2 target-genes. This is unexpected because the different DNMTs by and large seem to fulfil distinct roles in development and cellular viability². It is possible that various EZH2-containing complexes might exist that can interact with the *de novo* DNMTs or the maintenance DNMT under different circumstances. This would allow PcG complexes to take part in initiating and maintaining transcriptional programmes depending on the chromatin environment and stimuli.

Viré *et al.* show that overexpression of EZH2 increases the DNA methylation of the promoters of its target genes, and that, remarkably, when EZH2 is downregulated the methylation

decreases rapidly. This highlights the dynamic nature of DNA methylation and suggests that, just as histones are associated with separate enzymes for methylation and demethylation of their tails⁶, DNA demethylases may exist to actively remove the methyl marks, although such enzymes remain elusive.

How is DNA methyltransferase activity of DNMTs regulated by EZH2 histone methyltransferase activity? The authors propose a model whereby DNMTs require the EZH2-mediated histone H3-K27 methylation to methylate the DNA at target-gene promoters. This is analogous to the link proposed between the H3-K9 histone methylation system and DNA methylation in certain silent chromatin regions⁷. One possibility is that DNMTs recognize the H3-K27 methylation marks directly or indirectly through the interaction with another yet unknown partner that can regulate DNMT accordingly. Another possibility is that the H3-K27 methylation mark is not the engaging signal, but that DNMTs could be activated by EZH2 directly, perhaps by methylation.

DNMT recruitment to EZH2 target-genes not only adds another layer of repression to 'lock in' the silent state by methylating the surrounding DNA, but it might also help to recruit PRC1 proteins⁸. It is also possible that EZH2 acts independently of PRC1 by recruiting co-repressors that associate with DNMTs. In both scenarios, DNA methylation of EZH2 target-genes can lead to the recruitment of proteins that bind to the methyl-DNA marks, and the formation of more repressive complexes (Fig. 1). So are all EZH2 target-genes also silenced by DNA methylation, and are they overlapping with or separate from genes repressed by the EZH2-PRC1 interaction? Notably, the mechanism by which PcG-induced silencing is inherited is still not known. Perhaps, for a select group of target genes, EZH2-coupled DNA methylation can participate in preserving PcG-related cellular memory during DNA replication, through the maintenance DNMT (Fig. 2).

Although Viré *et al.* provide a simple and attractive model, their findings cannot easily be generalized to other PcG target-genes. For example, during X-chromosome inactivation, when one entire X chromosome is silenced in female mammals, EZH2 and another PcG protein known as EED accumulate on the inactive X chromosome at the onset of the process — well before there is any DNA methylation⁹. Moreover, paternal-specific repression of the *Kcnq1* gene domain in

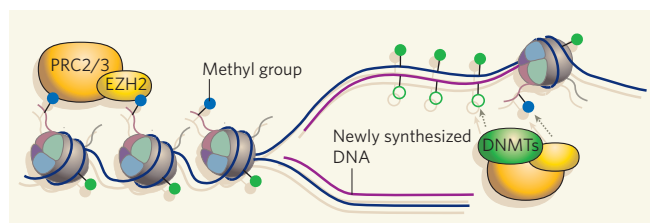


Figure 2 | Inheritance of epigenetic memory. A possible mechanism for inheritance of Polycomb-related epigenetic memory through DNA methylation during DNA replication. As the DNA is replicated, the maintenance DNA methyltransferase, DNMT1, adds a methyl mark (open green circles) to the newly replicated strand (purple), thus passing the DNA methylation mark to the progeny of the dividing cell through DNA replication. Viré *et al.*¹ report that DNMT1 interacts with the histone methyltransferase EZH2, contained in the PRC2/3 complexes, possibly recruiting the complexes to sites of newly methylated DNA, to ensure that the histones too are in a repressive state.

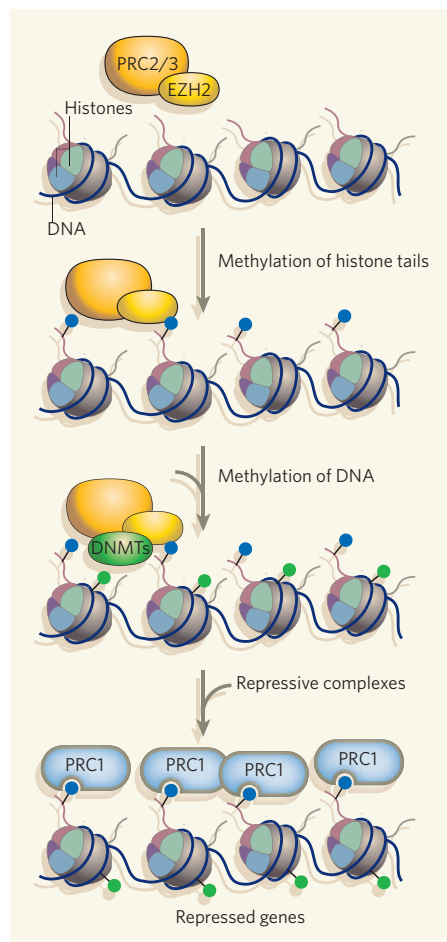


Figure 1 | Making epigenetic methylation marks. The Polycomb group protein EZH2 acts as part of the Polycomb repressive complexes 2 and 3 (PRC2/3), which add methyl groups to histone proteins, primarily to lysine 27 of histone H3 (small blue circles). Viré *et al.*¹ show that EZH2 interacts with DNA methyltransferases (DNMTs) to recruit them to the regulatory region of certain target genes. The DNMTs methylate DNA (small green circles), either setting up new methyl marks during development, or maintaining the marks after each round of cell division. These marks are correlated with gene repression, and they could have a role in recruiting further repressive complexes, such as PRC1 or histone deacetylase co-repressors, through methyl-binding domain proteins.

mouse placenta depends mainly on recruitment of EZH2 and EED with no apparent promoter DNA methylation^{10,11}. These and other examples^{12,13} imply that although PcG-related silencing and DNA methylation cooperate in many cases, they are not always coupled. One intriguing possibility is that DNMTs could be inactive when they are recruited by EZH2, only being activated later by some as yet unknown regulatory signal.

Finally, the core members of the PRC2 complex are evolutionarily conserved from plants to mammals³. Yet, the nematode *Caenorhabditis elegans* seems to lack any detectable DNA methylation or genomic sequences resembling genes encoding DNMTs¹⁴, and fly DNA has very little methylation^{5,15}. So, the connection between PcG proteins and DNA methylation must have evolved relatively recently.

Notwithstanding the questions that remain, the demonstration that the PcG protein EZH2 and DNA methylation can be directly coupled will provoke new hypotheses regarding the dynamics and heritability of epigenetic marks that can now be explored.

Panthea Taghavi and Maarten van Lohuizen are in the Division of Molecular Genetics, Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, the Netherlands. e-mail: m.v.lohuizen@nki.nl

- Viré, E. *et al.* *Nature* **439**, 871–874 (2006).
- Li, E. *Nature Rev. Genet.* **3**, 662–673 (2002).
- Kuzmichev, A., Jenuwein, T., Tempst, P. & Reinberg, D. *Mol. Cell* **14**, 183–193 (2004).
- Lund, A. H. & van Lohuizen, M. *Curr. Opin. Cell Biol.* **16**, 239–246 (2004).
- Ringrose, L. & Paro, R. *Annu. Rev. Genet.* **25**, 413–434 (2004).
- Yu-ichi, T. *et al.* *Nature* advance online publication doi:10.1038/nature04433 (2005).
- Lehnertz, B. *et al.* *Curr. Biol.* **13**, 1192–1200 (2003).
- Hernández-Muñoz, I., Taghavi, P., Kuijil, C., Neefjes, J. & van Lohuizen, M. *Mol. Cell. Biol.* **25**, 11047–11058 (2005).
- Heard, E. *Curr. Opin. Cell Biol.* **16**, 1–9 (2004).
- Umlauf, D. *et al.* *Nature Genet.* **12**, 1296–1300 (2004).
- Lewis, A. *et al.* *Nature Genet.* **12**, 1291–1295 (2004).
- Yu-Vuei, L. *et al.* *Cancer Res.* **64**, 8184–8192 (2004).
- Feldman, N. *et al.* *Nature Cell Biol.* advance online publication doi:10.1038/ncb1353 (2006).
- Grace Goll, M. & Bestor, T. H. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
- Marhold, J., Kramer, K., Kremmer, E. & Lyko, F. *Development* **131**, 6033–6039 (2004).

CHEMISTRY

Volatile times for ionic liquids

Peter Wasserscheid

Ionic liquids are useful substances, but in certain applications their utility is limited because they are involatile — or so we thought. In fact, some ionic liquids can be distilled, and even thermally separated.

The unique physical and chemical properties of ionic liquids — molten salts characterized by melting points below 100 °C — can be used to enhance the efficiency of a wide range of electrochemical, analytical, synthetic and engineering processes¹. This has generated huge interest in these liquids, with their first applications in industrial organic synthesis now being reported^{2,3}. On page 831 of this issue⁴, Earle and colleagues, a collaboration of researchers from Portugal, Northern Ireland and the United States, report that some ionic liquids are volatile, and so can be distilled. The findings shake up the received wisdom that ionic liquids are involatile as a class, and might extend their range of applications still further.

Exactly how an ionic liquid behaves varies sharply according to its specific ionic composition: for example, both highly hydrophobic and highly hydrophilic liquids can be produced (Fig. 1). Certain ionic liquids are known to undergo thermal decomposition through mechanisms such as the transfer of an alkyl group⁵ or, in the case of protic ionic liquids⁶ (whose cation is a proton-donating, or Brønsted, acid), through deprotonation. For some imidazolium-based ionic liquids, in which a weakly acidic proton is bound to a carbon atom, the formation of highly reactive organic molecules known as carbenes in the presence of a strong base has also been reported^{7,8}. The decomposition products are usually volatile, and, in some cases, reformation of an ionic liquid in colder parts of the apparatus has been observed.

Involatility had been assumed to be a property common to all ionic liquids that do not undergo thermal decomposition. Earle and colleagues⁴ change that assumption, showing that certain ionic-liquid structures known for their very high thermal stability, in particular bis{(trifluoromethyl)sulphonyl}amide salts, can be evaporated and recondensed under relatively mild conditions. The distillation of 1-hexyl-3-methylimidazolium bis{(trifluoromethyl)sulphonyl}amide, for instance, is reported to occur at 170 °C and at a pressure of 0.07 millibar.

It should be noted that Earle and colleagues do not give direct proof for the presence of ions in the gas phase of their distillation



Figure 1 | A class apart. Ionic liquids show a range of properties. Here, hydrophobic 1-ethyl-3-methylimidazolium bis{(trifluoromethyl)sulphonyl}amide (bottom phase) is separated from the equally hydrophobic alkane nonane by blue-dyed water. Involatility was, however, assumed to be a characteristic common to all members of the class — until Earle and colleagues⁴ showed otherwise.

processes. But they do describe several instances of ionic liquid evaporation and recondensation for which alkyl- and proton-transfer processes leading to neutral volatiles are both highly unlikely. Most interestingly, they report the separation by distillation of a mixture of two ionic liquids. This is also by itself no absolute proof that ions are present in the gas phase: the separation could, in at least one of the demonstrated cases, just reflect the different volatility of carbenes formed in a process of decomposition. Yet despite these reservations, the evidence for the volatility of at least some ionic liquids is convincing.

At first sight, these findings⁴ seem to contradict previous studies that have reported applications of ionic liquids reliant on their blanket non-volatility. Ionic liquids have, for instance, been used in extractive distillation processes, and tested⁹ for their long-term stability over three months at 175 °C and 50 millibar. Those

tests showed no significant loss of the ionic liquid. Ionic liquids have also been used in ultra-high vacuum conditions of 10^{-9} millibar for X-ray photoelectron spectroscopy analysis without decomposing or evaporating¹⁰.

Detailed analysis, however, reveals that the ionic liquids and reaction conditions used in these previous studies differ significantly from those used by Earle and colleagues. The clear corollary of the authors' findings is that non-volatility is not a property that can be assumed for all thermally stable ionic liquids: vapour pressure is just another physical-chemical property that depends on an ionic liquid's cation-anion combination, and will in the future have to be checked experimentally for each new liquid. Adjusting the vapour pressure of the ionic liquid will become an integral part of finding the best ionic liquid for a given application.

Do these developments bring into question the view that ionic liquids are 'green solvents'? That reputation is built largely on their non-volatility, because they do not create the atmospheric pollution that can result from the volatility of classic organic solvents. The answer is no: Earle and colleagues clearly state that the vapour pressure of ionic liquids remains negligible at near ambient conditions, and many ionic liquids show no signs of distillation below the temperature of their thermal decomposition. What the proven volatility of some thermally stable ionic liquids under relatively mild (and close to technically realistic) conditions will most assuredly do, however, is make these substances more useful for us. New purification methods and reactions that use ionic liquids in the gas phase are territories that are now ripe to be explored. The possibility of separating two ionic liquids by distillation will, in particular, unlock the door to new routes towards the

manufacture of ionic liquids, and towards the regeneration of spent ionic liquids from technical processes. ■

Peter Wasserscheid is in the Department of Chemical Reaction Engineering, University of Erlangen, Egerlandstraße 3, 91058 Erlangen, Germany.

e-mail: wasserscheid@crt.cbi.uni-erlangen.de

1. Wasserscheid, P. & Welton, T. (eds) *Ionic Liquids in Synthesis* (Wiley-VCH, Weinheim, 2003).
2. Freemantle, M. *Chem. Eng. News* **81**, 9 (2003).
3. Maase, M. in *Multiphase Homogeneous Catalysis* (eds Cornils, B. et al.) 560–566 (Wiley-VCH, Weinheim, 2005).
4. Earle, M. J. et al. *Nature* **439**, 831–834 (2006).
5. Maase, M. World patent WO 05/068404 (2005).
6. Volland, M. et al. World patent WO 03/062251 (2003).
7. Earle, M. J. & Seddon, K. R. World patent WO 01/077081 (2001).
8. Maase, M. & Massonne, K. World patent WO 05/019183 (2005).
9. Beste, Y., Eggersmann, M. & Schoenmakers, H. *Chem. Ing. Tech.* **77**, 1800–1808 (2005).
10. Smith, E. F., Villar Garcia, I. J., Briggs, D. & Licence P. *Chem. Commun.* 5633–5635 (2005).

P. WASSERSCHIED

GENETICS

Copies count

Joseph H. Nadeau and Charles Lee

Some genes have more than one copy, and the copy number can differ among individuals. But does this variation affect the person involved? It seems susceptibility to certain common diseases can be altered.

Each of our kidneys consists of about one million glomeruli — clusters of tiny blood vessels that together filter a remarkable 50 gallons of blood each day. Filtering separates waste products for excretion while retaining water and other vital substances such as glucose and proteins. In diseases such as diabetes and systemic lupus erythematosus (SLE), inflammation causes progressive loss of filtering abilities, leading to a condition known as glomerulonephritis that often results in kidney failure and renal disease. As with any disease caused by a combination of environmental and multiple genetic factors, tracking down the individual genetic components of glomerulonephritis has been a daunting task. In this issue, Aitman *et al.*¹ (page 851) provide compelling evidence that differences in the number and nature of a gene called *Fcgr3* contribute to the development and progression of the disease in rat models and humans. Their discovery suggests that copy-number variation, which is a major source of genetic variation in humans and other species, may contribute to susceptibility to other common diseases.

Despite the complications that result from the multigenic control of disease susceptibilities, the genes that contribute to common diseases such as heart disease are beginning to be identified². Examples in humans include the genes that contribute to asthma³, and to myocardial infarction and stroke⁴. In mice, certain genes have been implicated in age-related sensorineural hearing loss⁵ and susceptibility to atherosclerosis⁶. And several rat genes have been linked to unresponsiveness to insulin and to diabetes⁷. The *Fcgr3* gene is the most recent addition to the list of identified complex trait genes. Genes typically affect an individual's susceptibility to a disease because mutations change either the amount or the composition of the protein encoded by the gene. But in the case of *Fcgr3* in glomerulonephritis, the underlying molecular alteration is simply a change in the number of copies of the gene.

Aitman and colleagues' discovery¹ came from work on the Wistar Kyoto (WKY) rat strain, which is a model for glomerulonephritis in humans. Ten days after injection with a toxic serum⁸, these rats develop nephrotoxic nephritis, which has many pathological similarities to certain forms of human glomerulonephritis. By contrast, other rat strains such as the Lewis and Brown Norway strains do not succumb to the disease after the injection.

By breeding susceptible strains with

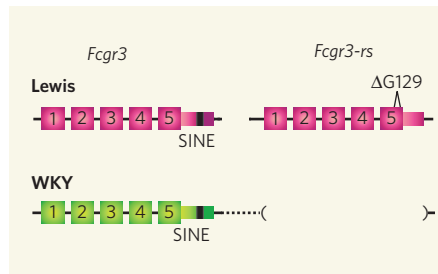


Figure 1 | Organization of the *Fcgr3* genes in Lewis and WKY rats. Chromosome 13 in Lewis rats has one copy of the *Fcgr3* gene and one copy of the related gene *Fcgr3-rs*. In WKY rats, the same chromosome contains only the one *Fcgr3* gene. *Fcgr3* in both strains has an extra 226 base pairs, including a so-called SINE element. The *Fcgr3-rs* gene in Lewis rats has a single nucleotide deleted (ΔG129). Aitman *et al.*¹ show that loss of the *Fcgr3-rs* gene in WKY rats contributes to susceptibility to nephrotoxic nephritis, a disease that recapitulates certain types of human glomerulonephritis.

resistant ones, Aitman *et al.* surveyed the rat genome for genes that are associated with nephrotoxic nephritis, and identified two important regions — one on chromosome 13 and the other on chromosome 16. The one on chromosome 13 is near a family of genes encoding receptors involved in immune responses, including *Fcgr3*. The authors found that WKY rats and other susceptible strains have only a single copy of the *Fcgr3* gene. But resistant strains including Lewis rats have a copy of the *Fcgr3* gene and a copy of the *Fcgr3*-related sequence (*Fcgr3-rs*), which probably evolved from a duplication of the *Fcgr3* gene (Fig. 1).

The *Fcgr3* and *Fcgr3-rs* genes in the rat have two differences. First, *Fcgr3* has an extra 226 base pairs inserted into the so-called 3'-untranslated region, which may regulate the amount of encoded protein expressed but doesn't affect the protein itself. The second difference involves the loss of a single nucleotide in the *Fcgr3-rs* gene (ΔG129) that alters the composition of the encoded receptor protein: the deletion results in a cytoplasmic domain in the protein that is longer by six amino acids.

The *Fcgr3* gene encodes a transmembrane receptor that is found on certain immune cells, including macrophages, and is involved in the inflammatory response. Activation of the receptor causes the immune cells to kill neighbouring non-immune cells, either by engulfing them (phagocytosis) or by

producing antibodies that kill them (cytotoxicity). Macrophages from the WKY rats produced greater cytotoxicity than macrophages from the Lewis rats. Moreover, adding *Fcgr3*, *Fcgr3-rs* or a variant of *Fcgr3* with the ΔG129 deletion to cells in culture can alter their behaviour. Cells with *Fcgr3* alone had higher levels of phagocytosis than those with *Fcgr3-rs* or *Fcgr3*-ΔG129. However, adding *Fcgr3-rs* to the *Fcgr3* cells inhibited phagocytosis, implying that *Fcgr3-rs* damps down the effects of *Fcgr3*. So when *Fcgr3* is lost in WKY rats, the macrophages become hyperactive. Together, the genetic and functional evidence satisfies the requirements of proof needed to establish that a genetic variant contributes to a complex multifactor trait².

But how does this relate to humans? Aitman *et al.* examined the *FCGR3B* gene (the human relative of *Fcgr3*) in individuals with glomerulonephritis associated with SLE. The inheritance patterns of this gene in some families affected by SLE had suggested that copy-number variation might underlie the condition. The authors found that from none to four copies of *FCGR3B* can exist in a cell, but this number is lower among SLE patients with glomerulonephritis than in unaffected control individuals. This indicates that reduced *FCGR3B* copy number is a risk factor for glomerulonephritis in SLE patients.

Copy-number variation is a major type of genetic variation among humans^{9,10}, and among inbred strains of mice^{11,12}. Currently, 1,237 copy-number variations have been identified¹³. Many of these occur in regions with large repeated or duplicated segments, which may increase the propensity for the DNA rearrangements that cause copy-number variations¹⁴. The functional attributes of proteins encoded by number-variant genes show that there is an enrichment for genes involved in drug detoxification, the immune response and inflammation, cell-surface integrity and cell-surface antigens¹⁵.

So, copy-number variants may well contribute significantly to the inter-individual variation seen in responses to drugs, immune defence and susceptibility to certain diseases. These variants could enhance or ameliorate many disease symptoms — as in the case of *Fcgr3* and glomerulonephritis, which is an often devastating complication of several common diseases. It is now possible to explore the functions of *Fcgr3* in the kidney, and perhaps to identify the naturally occurring agents that elicit the immune response leading to compromised kidney function in *Fcgr3* variants. Equally, it will be exciting to see what future discoveries will associate gene-copy-number variants with other common human diseases and their animal models¹⁶.

Joseph H. Nadeau is in the Department of Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA.
e-mail: jhn4@case.edu

Charles Lee is in the Division of Cytogenetics, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, 20 Shattuck Street, Boston, Massachusetts 02115, USA.
e-mail: clee@rics.bwh.harvard.edu

1. Aitman, T. J. *et al. Nature* **439**, 851–855 (2006).
2. Glazier, A. M., Nadeau, J. H. & Aitman, T. J. *Science* **298**, 2345–2349 (2002).
3. Zhang, Y. *et al. Nature Genet.* **34**, 181–186 (2003).
4. Helgadóttir, A. *et al. Nature Genet.* **36**, 233–239 (2004).
5. Noben-Trauth, K. *et al. Nature Genet.* **35**, 108–113 (2003).

6. Mehrabian, M. *et al. Circ. Res.* **91**, 120–126 (2002).
7. Aitman, T. J. *et al. Nature Genet.* **21**, 76–83 (1999).
8. Masugi, M. *Beitr. Pathol. Anat.* **91**, 82 (1933).
9. Iafrate, A. J. *et al. Nature Genet.* **36**, 949–951 (2004).
10. Sebat, J. *et al. Science* **305**, 525–528 (2004).
11. Li, J. *et al. Nature Genet.* **36**, 952–954 (2004).
12. Adams, D. J. *et al. Nature Genet.* **37**, 532–536 (2005).
13. <http://projects.tcag.ca/variation>
14. Samonte, R. V. & Eichler, E. E. *Nature Rev. Genet.* **3**, 65–72 (2002).
15. Feuk, L., Carson, A. R. & Scherer, S. W. *Nature Rev. Genet.* **7**, 85–97 (2006).
16. Check, E. *Nature* **437**, 1084–1086 (2005).

SPECTROSCOPY

NMR down to Earth

Janez Stepišnik

High-precision nuclear magnetic resonance spectroscopy generally requires the use of powerful magnets. But using Earth's magnetic field allows us to gain some of the same information on the cheap.

Are expensive superconducting magnets necessary to perform high-resolution nuclear magnetic resonance (NMR) spectroscopy? Not absolutely, say Stephan Appelt and his colleagues in the February issue of *Nature Physics*¹. They use a far less pricey source of magnetism — Earth's own magnetic field — to distinguish the chemical structures of various molecules containing hydrogen, lithium and fluorine. The level of accuracy they achieve is an order of magnitude better than that possible with the most advanced superconducting magnets.

When an external magnetic field is applied to an atomic nucleus, it induces a polarization in the direction of the intrinsic rotation, or 'spin', of that nucleus' constituent protons and neutrons. These spins align either parallel or antiparallel to the field, causing the quantum-mechanically allowed energy levels of the nucleus to split. At a frequency corresponding exactly to the difference between these energy levels, the nucleus can absorb electromagnetic radiation: the phenomenon known as nuclear magnetic resonance or NMR, first observed by Felix Bloch and Edward Mills Purcell in 1946 (for which achievement they won the Nobel Prize in Physics in 1952).

Soon after the initial discovery, it became clear that the effective magnetic field on a nucleus — and consequently the observed NMR frequency — is subtly changed by the effects of both orbiting electrons (the 'chemical shift')² and the spins of neighbouring nuclei ('J-coupling')³. This was the beginning of the triumphant success of NMR as a spectroscopic tool for exploring the composition and chemical environment of molecules in the liquid state. In the decades since, the need for higher sensitivity and lower spectral dispersion has demanded higher, more homogeneous magnetic fields, fuelling the development of powerful superconducting magnets.

Nowadays, the highest-resolution NMR techniques, using magnets producing field strengths of between 1 and 10 tesla, reproduce the hydrogen spectrum with a broadening of spectral lines caused by the instrumentation of less than a tenth of a hertz.

Compared with the fields that can be attained with superconducting magnets, Earth's magnetic field is weak: it varies from about 25 microtesla (μT) at the Equator to 75 μT at the poles, with geomagnetic field lines inclined, in Europe and North America, at an angle of about 60° to the (horizontal) surface. The field is not constant: currents in the ionosphere and disturbances from Earth's interior produce slow daily variations in the field with amplitudes of some 25 nanotesla (nT), and superimposed on these are further oscillations with periods of a few seconds and amplitudes of about 1 nT. Far enough from electric installations and other sources of artificial magnetic perturbation, however, proper shielding can reduce these shorter variations to about 0.1 nT s^{-1} , and local spatial gradients to below 1 nT m^{-1} . These variations are comparable to those found in the fields of artificial magnets.

The first observation of a nuclear magnetic effect in Earth's magnetic field — the free precession of proton spin⁴, akin to the precession of a spinning top in Earth's gravitational field — came not long after the discovery of NMR. The weakness of the geomagnetic field is such that it causes only a slight natural polarization in proton and neutron spins. Before an NMR measurement in the Earth field can be made, therefore, these spins generally have to be polarized by a high magnetic field, or polarization has to be transferred from more-readily polarizable electrons using a method known as dynamic nuclear polarization. Such techniques have been used for measurements of the precise 'Larmor' precession frequency for protons, of



50 YEARS AGO

...you reveal more than I think you were aware of when writing "a nation using...a minority language cannot escape bilingualism if it desires to attain high standards of scholarship". If the sentence is understood as referring to the use of minor languages for publishing, it is indisputable; but all too frequently it turns the other way — scholars within the major language groups neglecting the literature outside their own language... To illustrate this point I have made a small survey of world scientific literature... That our Soviet colleagues know more about 'Western' literature than the reverse is nothing new, but it is deplorable... That the English and even more the American literature should emerge as the narrowest is scarcely unexpected... It is a waste, and it is also inconsiderate, to publish primary scientific material in a minor language.

From *Nature* 18 February 1956.

100 YEARS AGO

Dr. H. Charlton Bastian re-expounds his well known biological heresies with a vigour and industry worthy of a better cause. The first heresy is that "archebiosis" is a present occurrence, that living organisms may here and now arise from non-living materials... we are recommended to take an infusion of turnip or fresh beef, to filter this through two layers of the finest Swedish paper, to let a drop fall on a cleaned microscope slip, to put a cover-glass on, to remove excess of fluid with blotting paper, to allow one or more air bubbles to remain in the film, to seal up with melted paraffin wax...to incubate at blood-heat for two to three hours, and to await events. The expected happens — multitudes of living particles appear... While we must stand aloof from Dr. Bastian's heresies, we cannot but admire his dogged support of what seems to us a lost cause. It is something to stand *unus contra mundum* with no loss of courage or good humour.

From *Nature* 15 February 1906.

50 & 100 YEARS AGO

the relaxation time that spins require to return to their normal states following polarization, and of the strength of J-coupling between nuclei⁵. The method has also been applied to the magnetic resonance imaging of pieces of fruit and of phantoms standing in for human tissue⁶, and for the detection of groundwater reservoirs⁷.

The disadvantage of low-field NMR for chemical analysis is that the chemical shift of spectral lines is smaller than the broadness

that is due to the magnetic field. This renders effectively unobservable the direct information about a nucleus' chemical environment gained from the chemical shifts caused by the surrounding electrons' screening of the applied magnetic field. On the other hand, the J-coupling between nuclear spins caused by chemical bonds is almost independent of the magnetic field, so its effect is proportionately greater where the field is low. In J-coupling, the quantum states of shared electrons impose

information about the chemical environment through the indirect dipolar coupling between nuclei that are relatively close to one another. The effect of the resultant J-splitting on the NMR frequency is at an observable level of between a few tenths and a few tens of hertz.

In their experiments, Appelt *et al.*¹ measure the J-coupled spectra of various compounds containing hydrogen, lithium and fluorine. They first pre-polarized the nuclei of each sample in the field of a permanent magnet to a



CARTOGRAPHY

A popular perspective

Maps shape our perception of geographical realities, albeit often imperfectly. A world map, for instance, relies on some form of projection that transfers — with inevitable distortion — the unbounded surface of the globe to the bounded page. Mark Newman has taken this idea of distortion a stage further, producing a world map that gives each territory a size proportional to its population.

The idea of such density-equalized 'cartograms' is not new, and can be applied to a host of variables, besides population, whose distribution is geographically uneven. The innovation in Newman's map (pictured) is the use of the diffusion equation, familiar from the physics of heat transfer and molecular

mixing, to produce a map that smooths out population density, but keeps geographical distortion to a minimum.

The map follows on from work in which Newman and Michael T. Gastner applied the diffusion technique to (among other things) maps depicting the results of the 2000 US presidential election (*Proc. Natl Acad. Sci. USA* **101**, 7499–7504; 2004). For the world cartogram, a grid of 4,096 by 2,048 squares was overlaid on a rectangular world map based on a cylindrical equal-distance (plate carrée) projection. A starting population-density function was computed by dividing the population of each country equally between the squares covering its territory.

Population was then allowed to diffuse away from areas of higher

density to those of lower density, with national boundaries moving such that the net population flow through them was zero at all times. The displacement of the boundaries was recalculated throughout the diffusion process by integrating the diffusion-velocity field, until the population density was equalized over the land surfaces. Constituent parts of non-contiguous territories (such as nations divided between several islands) were treated as separate entities, and the oceans and unpopulated Antarctica were treated as a sea of uniform population density equal to the average density of the populated land areas. This ensured an end result that maintained the semblance of familiar geographical relationships.

Apart from the domination of Asian nations — China and India alone account for about a third of the world population of some 6.6 billion — the map throws up many other

interesting facets. The Americas are diminished; Russia and Canada, the two largest geographical territories, are reduced to Arctic buffer zones. Australia all but disappears, especially when compared with its near neighbour Indonesia.

The value of Gastner and Newman's technique lies in its intuitiveness and the relative simplicity of the algorithm, which requires a computation time of minutes. The diffusion method may also be extended readily to three dimensions, to create, for example, a three-dimensional 'homunculus' in which each part of the body is scaled according to the size of the brain region devoted to it. The potential of such techniques for producing startling representations that challenge our preconceptions seems — unlike world maps — unbounded.

Richard Webb

value about 10,000 times that attainable in the geomagnetic field, and transferred the sample within one second into the detection coil of an Earth-field NMR spectrometer. They investigated benzene, lithium chloride dissolved in water, tetramethylsilane, silicone oil, octamethylcyclotetrasiloxane and nonafluorohexene, demonstrating in each case that their spectrometer could image these compounds' molecular structure. The instrumental broadening of their spectrometer is only a few millihertz over a sample volume of 1 cm³ — less than the broadening of high-resolution NMR with advanced superconducting magnets. The spectral resolution is, however, limited by the shorter spin relaxation time to a few tenths of a hertz.

The authors predict that the combination of low-field NMR with advanced methods of spin hyperpolarization will allow time-resolved NMR spectroscopy of rare nuclei such as ⁶Li, ¹³C and ²⁹Si, and the separation of the corresponding J-couplings with high precision. Many applications of this mobile low-field NMR technique are also likely in medicine and materials science: the tracking of hyperpolarized ⁶Li⁺ and ⁷Li⁺ migrating through ion channels or membranes, for example, or the characterization of mineral oil in well-logging and the online detection of chemical reactions.

Earth-field NMR cannot replace the diagnostic capabilities of the chemical shift measurements available with the high-field method. But particularly in combination with

superconducting quantum interference device (SQUID) detectors⁸, the technique could give limited, but potentially very useful, information about the coupling of many molecules.

Janez Stepišnik is in the Physics Department, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia.

e-mail: Janez.Stepisnik@fmf.uni-lj.si

1. Appelt, S., Kühn, H., Wolfgang, F. & Blümich, B. *Nature Phys.* **2**, 105–109 (2006).
2. Proctor, W. G. & Yu, F. C. *Phys. Rev.* **77**, 717 (1950).
3. Ramsey, N. F. & Purcell, E. M. *Phys. Rev.* **85**, 143–144 (1952).
4. Waters, G. S. & Francis, P. D. *Sci. Instrum.* **35**, 88–93 (1958).
5. Bene, G. J. *Phys. Rep.* **58**, 213–267 (1980).
6. Stepišnik, J., Eržen, V. & Kos, M. *Magn. Reson. Med.* **15**, 386–391 (1990).
7. Shushakov, O. A. *Magn. Reson. Imaging* **14**, 959–960 (1996).
8. McDermott, R. et al. *Science* **295**, 2247–2249 (2002).

CHEMISTRY

Brevity for bonds

Jay S. Siegel

Fashion design? Game-playing? Designing the shortest bond between two carbon atoms can seem to have elements of such apparently ephemeral pursuits. But it can also stretch the chemist's creativity.

Carbon–carbon (C–C) bonds could be tailored shorter this coming season, according to Deborah Huntley and colleagues, who include the Nobel laureate Roald Hoffmann. Their predictions appear in an unlikely fashion magazine, the journal *Angewandte Chemie International Edition*¹. The authors highlight several ways to sport short bonds, but it will be no small task for manufacturers to bring these designs into reality.

Molecular fashion has long been the chemist's playground² — no wonder, given chemistry's unique position in the design and synthesis of new forms of matter. The pursuit of novel molecular structures has led chemists to uncover material properties that cannot be revealed by observational science alone. Such creativity blends the hard science of chemistry with the engineering and aesthetic components of architecture and sculpture — or even *haute couture*. A question that often stems from a molecular-design hypothesis is “Can we make that?”. Whether the answer is yes or no, addressing such questions in a scholarly manner can generate great chemical insight.

In their paper, Huntley and colleagues¹ ask how many ways a C–C bond can be squeezed, and what is the resultant degree of shortening. They cite precedent from Henning Hopf's 1976 essay “Wie kurz kann eine Einfachbindung werden? Wie stark lässt sie sich stauchen?” (“How short can a single bond be? How strongly can it be compressed?”)³. As such, their call for a renewed interest in short bonds can be seen as a revivalist trend, which could signal a return to training chemists in

their basic trade: the synthesis and characterization of new molecular entities.

The authors propose three basic approaches to squeezing C–C bonds: caging, crowding and clamping. They evaluate all three computationally, using a reasonable computational level for the rough screening of structures. Such simulations represent a departure by

Hoffmann from his normally more poetic approach to theory. But given the goal of designing the shortest bond, one is left to grind out the numbers, or grind out the synthesis and structure determination.

In the end, investigating the concept of relative molecular compressibility teaches us more about the behaviour of molecules than can the size of the deviation from the ‘normal’ bond length of 1.53 Å. Huntley *et al.* predict a world record of 1.32 Å, but there is a chemistry lesson to be learnt from any compound that bears a C–C single bond less than 1.4 Å long, for example.

The shape of the ‘potential-energy well’ for stretching a C–C bond — the curve showing the relationship between the system's potential energy and the distance between two carbon atoms — is asymmetric; distortions to longer

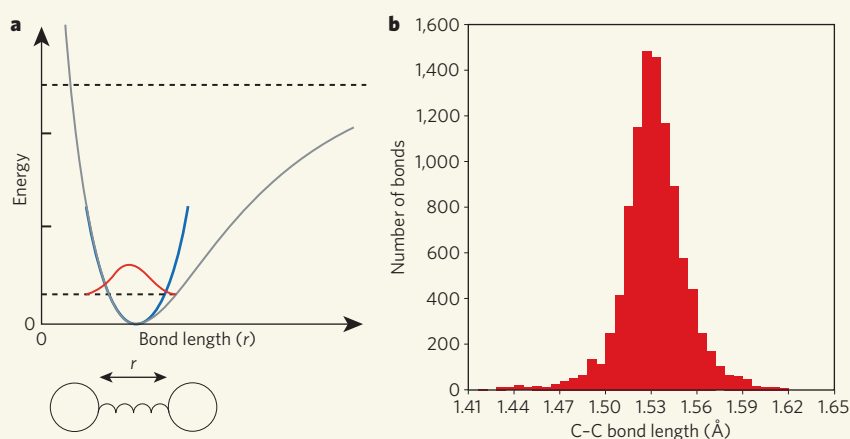


Figure 1 | Bonding energy potential and carbon–carbon bond-length distribution. **a**, The C–C bond can be considered energetically as two balls joined by a spring. Pulling them apart becomes increasingly harder (requires more energy) the longer the bond gets. Pushing them together will also become more difficult as they almost come into contact. But somewhere in the middle is a stable state. Although this approximation (the blue line) works fairly well, especially at the bottom of the energy well, a more realistic energy profile (grey line) is asymmetrical because the energy of stretching the bond is limited by the bond dissociation energy. The energy of interaction decreases as the atoms move apart; the red line shows a qualitative expectation of the distribution observed for **b**. **b**, Distribution of observed bond lengths, compiled from data from the Cambridge Structural Database⁵. Skew towards longer bonds becomes evident only past the 3σ level of deviation beyond the ‘normal’ bond length of 1.53 Å.

MICROBIOLOGY

Protection racket

When food gets scarce, cannibalism is always an option. In times of famine, soil-dwelling *Bacillus subtilis* bacteria (pictured) shut down, forming tough spores that can lie dormant for years. But before that, they stock up on nutrients by releasing a toxin that kills off their neighbouring siblings and then feed off the remains of the dead cells. But how do the toxin-producing bugs shield themselves from the poison's deadly effects? Craig Ellermeier *et al.* (*Cell* **124**, 549–559; 2006) have unravelled the surprisingly

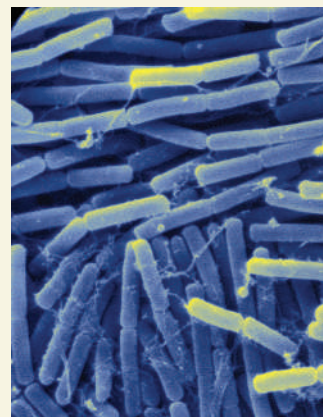
simple mechanism that protects the toxin-producers.

When nutrients are in short supply, a protein called SpoOA is activated to coordinate the cells' response. This regulator seems to be controlled by a bistable genetic switch such that it is turned on in only about half of the cells in a bacterial population. The activated SpoOA triggers manufacture of the toxin (SdpC) and an 'immunity' protein.

With some nifty genetics, Ellermeier *et al.* have identified the immunity protein and worked out how it is regulated. The key to

the process is the fact that the immunity protein is also a signal-transduction protein that controls its own synthesis. As toxin accumulates outside the cell, the immunity protein — which lies in the cell membrane — mops it up, preventing the cell from being killed.

At the same time, the toxin works with the immunity protein to grab hold of a repressor protein in the cell and tether it to the cell membrane. The repressor normally inhibits the gene that encodes the immunity protein, but it cannot function when bound to the membrane. So the cell keeps pumping out more immunity protein. When all the toxin is mopped up, free repressor starts to



S. LOWRY/UNIV. ILLUST/GETTY

accumulate, shutting down further synthesis of the immunity protein. Thus, the cell produces only as much immunity protein as it needs.

Helen Dell

distances cost less energetically than those to equivalently shorter ones. One might therefore think that longer bonds would form far more readily than shorter ones. Indeed, spectacularly long bonds of more than 1.7 Å have been reported⁴. However, the potential energy is virtually symmetric near the bottom of the potential well, like the potential-energy well of a simple ball-and-spring model of a molecule (Fig. 1a). Empirically (from 50,000 hits obtained from the Cambridge Structural Database⁵), observed C–C bond lengths are distributed with no apparent skew towards longer bonds well past the three-standard-deviation (3σ) mark (bond length $r = 1.53$ Å, $\sigma = 0.015$ Å; Fig. 1b). This result is consistent with Huntley and colleagues' computation that a compression of 0.05 Å would 'cost' only 0.7 kJ per mol more than elongation. To a chemist thinking of the distribution as being equilibrated at room temperature, such an energy difference would suggest a long-to-short ratio of 3:4 in the numbers of C–C bonds.

With such a wide range of 'naturally' occurring bonds, the design component becomes a novelty only at deviations beyond 6σ (about 0.1 Å), with especial interest focused on deviations of more than 10σ (0.15 Å) — bonds longer than 1.7 Å or shorter than 1.4 Å. At such extremes, the asymmetry of the potential-energy well does play a substantial role in the energetics, and this is also reflected in the bond-length distribution. In particular, experimentally sound C–C bond distances below 1.4 Å are big news.

Coincidentally, Huntley *et al.* estimate that the hypothetical energy required to compress a C–C bond by 10% (0.15 Å) is roughly 10% (38 kJ per mol) of the bond energy, if directed along the bonding axis. This energy requirement would essentially preclude the natural formation of structures with very short C–C bonds, and would relegate such bonds to the non-natural products pursued by those

engaged in chemical design and synthesis.

How might one go about designing a short C–C single bond? Let's begin by recognizing that an observed molecular structure is a structure in equilibrium, where the forces are balanced and potential energy is minimized. Distortion of a ball-and-spring model of a molecule reveals that molecules respond to stresses by distributing the energy into a variety of distortion modes, favouring those modes for which the forces can be balanced with minimum increase in internal energy⁶. This is sometimes presented as a sort of structural compensation principle, in which stretching one bond correlates with shrinking another, or the widening of one angle correlates with the narrowing of another. One example is that of octahedral distortions, in which the lengths of bonds to ligands in axial positions correspond inversely to those in the equatorial plane.

To make a C–C bond shorter, one needs to push, clamp or tug the carbons together while providing a scaffolding that is sufficiently stiff. Distances between non-bonded atoms, dihedral angles and bond angles all provide 'softer' distortion modes (that is, weaker spring potentials) than bond-length distortions to dissipate strain energy and weaken the compression. The use of axial symmetry can help to focus the compression along the bonding axis. Thus, the three-fold symmetric cages presented by Huntley and colleagues¹ make good design sense.

As a historical note, it should be remembered that the pursuit of molecules with compressed bonds started before the advent of routine X-ray and computational *ab initio* analysis. Examples are hexaisopropylbenzene⁷, the 'iron maidens' (cyclophanes) of Robert Pascal^{8,9} and intermediates en route to the 'bird-cage' hydrocarbons^{10,11}. All of these display large shifts in their carbon–hydrogen infrared stretching modes, a sign that they

possess compressed carbon–hydrogen bonds. In the case of bird-cage hydrocarbons, that bond length was found by neutron diffraction to be 1.078 Å, compared with the standard average length of 1.115 Å.

Coupled with chemical synthesis and structure elucidation, molecular design has long been an engine of creativity in chemistry. The notion of squeezing C–C bonds may seem pure game-playing, or a fashion that won't last beyond the current season. But it is simply one aspect of molecular design, and a myriad of problems involving extreme molecular geometries can be envisaged. For example, from the idea of a squeezed bond, one might try to create a molecular gauge in which the distortions of a common scaffold would be translated to an energy of compression for a variety of compressed central bonds. In any case, the Cambridge Structural Database⁵ is an invaluable reference for what is 'normal': from there, designing the abnormal may lead to a deeper understanding of chemical structure. ■

Jay S. Siegel is at the Organic Chemistry Institute, University of Zurich, Winterthurerstrasse 190, 8050 Zurich, Switzerland.
e-mail: jss@oci.unizh.ch

- Huntley, D. R., Markopoulos, G., Donovan, P. M., Scott, L. T. & Hoffmann, R. *Angew. Chem. Int. Edn* **44**, 7549–7553 (2005).
- Nickon, A. & Silversmith, E. F. *Organic Chemistry: The Name Game — Modern Coined Terms and their Origins* (Pergamon, Oxford, 1987).
- Hopf, H. *Chem. Unserer Zeit* **10**, 114–120 (1976).
- Baldridge, K. K., Ogawa, Y. K., Siegel, J. S., Tanaka, K. & Toda, F. *J. Am. Chem. Soc.* **120**, 6167–6168 (1998).
- www.ccdc.cam.ac.uk
- Bürgi, H.-B. & Dunitz, J. D. (eds) *Structure Correlation Vols 1, 2* (Wiley-VCH, Weinheim, 1994).
- Siegel, J., Gutierrez, A., Schweizer, W. B., Ermer, O. & Mislav, K. *J. Am. Chem. Soc.* **108**, 1569–1575 (1986).
- Pascal, R. A. Jr *Eur. J. Org. Chem.* 3763–3771 (2004).
- Song, Q., Ho, D. M. & Pascal, R. A. Jr *J. Am. Chem. Soc.* **127**, 11246–11247 (2005).
- Winstein, S., Carter, P., Anet, F. A. L. & Bourn, A. J. R. *J. Am. Chem. Soc.* **87**, 5247–5249 (1962).
- Ermer, O., Mason, S. A., Anet, F. A. L. & Miura, S. S. *J. Am. Chem. Soc.* **107**, 2330–2334 (1985).

BRIEF COMMUNICATIONS

Invasion and the evolution of speed in toads

Cane toads seem to have honed their dispersal ability to devastating effect over the generations.

Cane toads (*Bufo marinus*) are large anurans (weighing up to 2 kg) that were introduced to Australia 70 years ago to control insect pests in sugar-cane fields. But the result has been disastrous because the toads are toxic and highly invasive. Here we show that the annual rate of progress of the toad invasion front has increased about fivefold since the toads first arrived; we find that toads with longer legs can not only move faster and are the first to arrive in new areas, but also that those at the front have longer legs than toads in older (long-established) populations. The disaster looks set to turn into an ecological nightmare because of the negative effects invasive species can have on native ecosystems^{1,2}; over many generations, rates of invasion will be accelerated owing to rapid adaptive change in the invader³, with continual 'spatial selection' at the expanding front favouring traits that increase the toads' dispersal^{4,5}.

Introduced to Queensland in 1935, cane toads have since expanded their range to encompass more than a million square kilometres of tropical and subtropical Australia⁶. We have radio-tracked toads (for methods, see supplementary information) at the invasion front 60 km east of Darwin and confirmed the astonishing locomotor performance in these animals, which move up to 1.8 km per night during the rainy months — far further than previously studied anurans⁷. Does this remarkable ability result from selection for enhanced dispersal during the toads' Australian colonization history?

The morphological trait most often linked to locomotor ability in anurans is leg length, both among and within species⁸. Our trials (see supplementary information) confirm that cane toads with relatively long legs are indeed faster over a short distance (regressing time taken to cover 1 m against residual leg length: $r = -0.44$, $n = 29$, $P < 0.02$). But, more important, longer-legged toads moved further over 24 h (maximum displacement of radio-tracked toads versus relative leg length: $r = 0.46$, $n = 21$, $P < 0.04$) and over three days ($r = 0.58$,

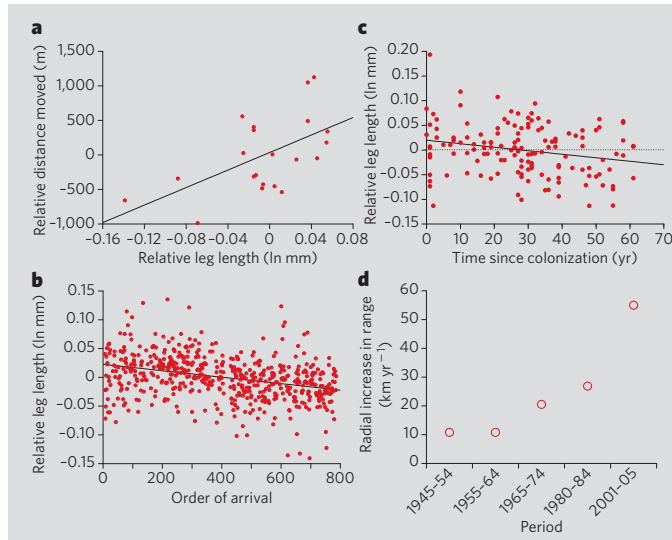


Figure 1 | Morphology of cane toads in relation to their speed and invasion history. **a, b,** Compared with their shorter-legged conspecifics, cane toads with longer hind limbs move further over 3-day periods ($r^2 = 0.34$) (**a**), and are in the vanguard of the invasion front (based on order of arrival at the study site; $r^2 = 0.11$) (**b**). **c,** Cane toads are relatively long-legged in recent populations, and show a significant decline in relative leg length with time in older populations ($r^2 = 0.05$). **d,** The rate at which the toad invasion has progressed through tropical Australia has increased substantially since toads were first introduced in 1935 ($r^2 = 0.92$).

$n = 21$, $P < 0.006$; Fig. 1a). Longer legs therefore facilitate more rapid dispersal.

If the invasion process has been assisted by the evolution of improved dispersal ability among toads at the front, three consequences would be expected. First, longer-legged toads should be disproportionately common among the first wave of arrivals at any site. As the toad invasion front passed our study site, we measured relative leg lengths of all toads encountered over a 10-month period. Longer-legged toads were the first to pass through, followed by shorter-legged conspecifics (order of arrival versus relative leg length: $r = -0.34$, $n = 552$, $P < 0.0001$; Fig. 1b). Longer-legged toads therefore moved faster through the landscape.

Second, toads at the invasion front should be longer-legged than toads from older populations. As predicted, longer-term historical analysis within Queensland populations shows that relative leg length is greatest in new arrivals and then declines over a 60-year period (Fig. 1c; $r = -0.23$, $n = 139$, $P < 0.008$).

Third, the rate of progress of the toad invasion front should increase through time. As predicted, rates of frontal progress have con-

sistently increased (Fig. 1d; time versus annual rate of spread, Pearson's $r = 0.96$, $P < 0.005$). Toads expanded their range by about 10 km a year during the 1940s to 1960s, but are now invading new areas at a rate of over 50 km a year. Accordingly, previous predictions about the time course of future expansion of the toads' range⁹ seriously underestimate their actual rates of movement.

These rapid shifts in toad morphology, locomotor speed and invasion velocity indicate that conservation biologists and managers need to consider the possibility of rapid adaptive change in invading organisms. If there is no fitness disadvantage to individual organisms at the invasion front, evolutionary forces are likely to fine-tune organismal traits in ways that facilitate more rapid expansion of the invading population¹⁰. Hence, control efforts against feral organisms should be launched as soon as possible, before the invader has had time to

evolve into a more dangerous adversary.

Benjamin L. Phillips, Gregory P. Brown, Jonathan K. Webb, Richard Shine

School of Biological Sciences A08, University of Sydney, New South Wales 2006, Australia
e-mail: rics@bio.usyd.edu.au

1. Crossland, M. R. *Ecography* **23**, 283–290 (2000).
2. Smith, K. G. *Biol. Conserv.* **123**, 433–441 (2005).
3. Cox, G. W. *Alien Species and Evolution* (Island, Washington, 2004).
4. Simmons, A. D. & Thomas, C. D. *Am. Nat.* **164**, 378–395 (2004).
5. Travis, J. M. J. & Dytham, C. *Evol. Ecol. Res.* **4**, 1119–1129 (2002).
6. Lever, C. *The Cane Toad. The History and Ecology of a Successful Colonist* (Westbury, Otley, West Yorkshire, 2001).
7. Smith, M. A. & Green, D. M. *Ecography* **28**, 110–128 (2005).
8. Choi, I., Shim, J. H. & Ricklefs, R. E. *J. Exp. Zool.* **299A**, 99–102 (2003).
9. Freeland, W. J. & Martin, K. C. *Aust. Wildl. Res.* **12**, 555–559 (1985).
10. Thomas, C. D. *et al. Nature* **411**, 577–581 (2001).

Supplementary information accompanies this communication on Nature's website.

Received 24 November 2005; accepted 24 January 2006.

Competing financial interests: declared none.
doi:10.1038/439803a

Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*

Ariel Goldraij^{1*}, Katsuhiko Kondo^{2*}, Christopher B. Lee^{2,3}, C. Nathan Hancock⁴, Mayandi Sivaguru³, Sonia Vazquez-Santana⁵, Sunran Kim², Thomas E. Phillips³, Felipe Cruz-Garcia⁶ & Bruce McClure²

Pollen-pistil interactions are crucial for controlling plant mating. For example, S-RNase-based self-incompatibility prevents inbreeding in diverse angiosperm species. S-RNases are thought to function as specific cytotoxins that inhibit pollen that has an S-haplotype that matches one of those in the pistil. Thus, pollen and pistil factors interact to prevent mating between closely related individuals. Other pistil factors, such as HT-B, 4936-factor and the 120 kDa glycoprotein, are also required for pollen rejection but do not contribute to S-haplotype-specificity *per se*. Here we show that S-RNase is taken up and sorted to a vacuolar compartment in the pollen tubes. Antibodies to the 120 kDa glycoprotein label the compartment membrane. When the pistil does not express HT-B or 4936-factor, S-RNase remains sequestered, unable to cause rejection. Similarly, in wild-type pistils, compatible pollen tubes degrade HT-B and sequester S-RNase. We suggest that S-RNase trafficking and the stability of HT-B are central to S-specific pollen rejection.

Plants have unique mate selection challenges because of their reliance upon external vectors such as wind or animals to distribute pollen. To meet these challenges, they use a variety of ethological, structural and biochemical mechanisms to control mating. Self-incompatibility (SI) systems prevent inbreeding by recognizing and rejecting self-pollen and pollen from closely related individuals¹. The most widespread SI system uses cytotoxic S-RNases secreted by the pistil to selectively inhibit self-pollen². This form of SI is genetically controlled by a polymorphic S-locus containing two genes that control S-specificity. The *S-RNase* gene controls specificity in the pistil^{3,4}. Cells of the stigma, style and ovary secrete S-RNase into the extracellular matrix (ECM) that guides the pollen tube to the ovule^{5,6}. The *S-locus F-box* gene (pollen S, *SLF*) is expressed in pollen and pollen tubes, and controls specificity in pollen^{7–12}. Other factors, which contribute to non-S-specific functions, are also required¹³. HT-B is a small asparagine-rich protein expressed at a late stage of style development¹⁴. Antisense- or RNA interference (RNAi)-mediated suppression of HT-B in *Nicotiana* and *Solanum* causes failure of S-specific pollen rejection without directly affecting S-RNase^{14,15}. Furthermore, changes in HT-B may have been responsible for mating-system shifts in *Solanum*^{16,17}. The 120 kDa glycoprotein (120K) is an abundant arabinogalactan protein in the stylar ECM that is taken up by growing pollen tubes^{18,19}. The 120K protein binds to S-RNase *in vitro* and, like HT-B, suppressing its expression by RNAi prevents S-specific pollen rejection^{20,21}. Other factors are known only from their genetic effects^{22–26}. We have studied mutations in *4936-factor* that result in self-compatibility in *Nicotiana*¹³. The corresponding gene has not been cloned, but it is known to act on only the style side and is distinct from the *HT-B* and *120K* genes.

Much of the current research focuses on how S-RNase interacts with SLF to cause S-specific pollen rejection. S-RNase enters both incompatible and compatible pollen tubes²⁷, but the cytotoxic effect of S-RNase, manifested by degradation of pollen RNA and growth inhibition, only occurs in incompatible pollen tubes²⁸. In the most popular models, this selectivity is explained by postulating that pollen S inhibits non-self S-RNase². The identification of pollen S as an F-box protein (that is, SLF), potentially involved in ubiquitin-mediated protein degradation, led to the hypothesis that inhibition results from selective degradation of non-self S-RNase^{9,11,29}. Results consistent with this model show that SLF from *SI Antirrhinum* interacts with S-RNase in a non-S-specific manner *in vitro* and in the yeast two-hybrid system²⁹. This model, however, does not address non-S-specific aspects of pollen rejection, such as S-RNase uptake into both compatible and incompatible pollen tubes²⁷. In addition, a self-compatible mutant in *Prunus avium* seems to have an *SLF* gene deletion, suggesting that SLF is not necessary for resistance to S-RNase in the Rosaceae³⁰. Finally, although it is known that non-S-specific factors are required for SI^{14,21}, they are not represented in current models^{11,12,29,31}. Here we report that S-RNase is compartmentalized in pollen tubes. In the absence of *HT-B* or *4936-factor*, it remains compartmentalized and cannot cause pollen rejection. Although levels of the S-RNase protein seem stable, HT-B is degraded in compatible pollen tubes. We present an alternative model to explain these results.

Style-side factors that act late in pollen rejection

We investigated S-RNase uptake in plants with defects in *HT-B* and *4936-factor*. The *HT-B*-antisense plant used here shows no detectable HT-B protein and fails to appropriately reject *S_{C10}*-

¹CIQUIBIC, Departamento de Química Biológica, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina. ²Division of Biochemistry, 105 Life Sciences Center, 1201 E. Rollins Street, Columbia, Missouri 65211, USA. ³Division of Biological Sciences, 2 Tucker Hall, Columbia, Missouri 65211, USA.

⁴Department of Plant Biology, 4505 Miller Plant Sciences Building, Athens, Georgia 30602, USA. ⁵Departamento de Biología, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ave Universidad y Copilco, México D.F. 04510, México. ⁶Departamento de Bioquímica, Facultad de Química, Universidad Nacional Autónoma de México, Ave Universidad y Copilco, México D.F. 04510, México.

*These authors contributed equally to this work.

pollen¹⁴. Similarly, in *Nicotiana alata* populations that segregate for 4936-factor mutations, 'rejectors' show normal S-specific pollen rejection while 'acceptors' fail to reject pollen appropriately. Figure 1 shows the phenotypes of the segregants used in this study. Both the wild-type rejector and the 4936-factor-mutant acceptor express HT-B, and S_{C10^-} and S_{105^-} -RNase (Supplementary Fig. S1). However, the rejector shows normal S-specific pollen rejection, but the acceptor is compatible with S_{105^-} and all other pollen S-haplotypes tested (Fig. 1a versus b). Thus, the phenotypes are similar; HT-B-antisense plants and 4936-factor-mutant plants fail to reject pollen appropriately, and do so without affecting S-RNase levels.

We used immunolocalization to determine whether HT-B and 4936-factor affect S-RNase uptake. Double-labelling experiments, using anti-S-RNase and anti-callose antibodies, show that the S-RNase distribution in pollen tubes growing in pistils of both defective backgrounds is similar to controls (Fig. 1c–f). Similar results were obtained in plants in which 120K expression was suppressed by RNAi (discussed below). Our results confirm previous studies²⁷ showing abundant S-RNase uptake in both compatible and incompatible pollen tubes; S-RNase labelling inside pollen tubes is frequently as intense, or more so, than labelling of the ECM. Although S-RNase labelling inside pollen tubes is often widespread, in some pollen tubes it is localized to specific areas (Fig. 1c–f, arrows).

The results suggest that these style-side factors act after S-RNase uptake but before it exerts its cytotoxic effect. Clearly, in these style-side defective plants, pollen tubes can accommodate large amounts

of S-RNase and compatibility is not necessarily due to large scale degradation of S-RNase. Further experiments addressed both why the S-RNase in style-side defective backgrounds cannot cause pollen rejection, and how non-S-specific factors may be implicated in compatibility.

S-RNase is compartmentalized in compatible pollen tubes

We recently showed that an S-RNase binding protein, 120K, is required for S-specific pollen rejection²¹. Therefore, we performed triple-labelling experiments using antibodies for callose, S-RNase and 120K to investigate colocalization of 120K and S-RNase in pollen tubes. The results revealed that S-RNase is sequestered in a pollen tube compartment marked by the 120K protein (Fig. 2). 120K sometimes overlaps with S-RNase in the pollen tube, but it also frequently appears as series of closely spaced points that define the boundary of a compartment containing S-RNase (Fig. 2a–d). Orthogonal sectioning of the confocal image stack shows the compartment surrounding S-RNase in three dimensions (Fig. 2a–d, f; y–z plane images). When the compartment assumes a ribbon-like appearance (for example, Fig. 2a–c, f), the compartmentalization of S-RNase is especially apparent. Ribbon-like vacuoles have been observed in the cytoplasmic tip regions of living *Arabidopsis thaliana* pollen tubes³². We believe that images of S-RNase-filled pollen tubes represent highly vacuolated regions distal to the tip, and that the regions with ribbon-like compartments are proximal to the tip. A previous study described S-RNase uptake, but suggested that it is not contained in a compartment²⁷; we reached the opposite conclusion. This discrepancy may be due to the earlier use of fixation conditions preserving antigenicity at the expense of visualizing membrane structures, and the lack of a marker for internal membranes²⁷.

We used vacuolar markers to further characterize the compartment. Aleurain is a luminal cysteine protease present in vacuoles of barley, petunia and tobacco³³. Aleurain antibody staining labels numerous small vesicles in transmitting tract cells (Fig. 2g). It also labels the lumen of the ribbon-like pollen tube compartment in a manner similar to the S-RNase antibody, although the labelling is not as strong (for example, Fig. 2a versus g). Vacuolar pyrophosphatase (vPPase) is a marker for the vacuolar membrane³⁴. An antibody to vPPase labels transmitting tract cell vacuoles as well as the ribbon-like compartment in compatible pollen tubes (Fig. 2h). It also labels the compartment in 120K-RNAi plants, where the 120 kDa glycoprotein has been suppressed below the limit of detection²¹, and in the HT-B-antisense plant (Supplementary Fig. S2).

The immunolocalization results suggest that S-RNase and 120K are associated with the pollen tube endomembrane system. We tested this further by dissecting pollen tubes from styles heavily pollinated with compatible or incompatible pollen, and analysing membrane fractions. The pollen tube preparations are enriched for pollen tubes, but also contain transmitting tract cells and ECM material (Supplementary Fig. S3a). After high-speed centrifugation, vPPase pelleted with portions of the S-RNase, HT-B and 120K (10,000g supernatant (S10) and 156,000g supernatant (S156) versus 156,000g pellet (P156); Supplementary Fig. S3b). Much less HT-B was recovered from compatible pollen tubes. The results showing that most of the S-RNase and 120K was contained in the S156 fraction were expected, because these proteins are abundant and highly soluble components of the ECM. In addition, most of the luminal contents from large vacuoles are released upon homogenization, and only proteins immediately adjacent to the membrane are pelleted.

S-RNase uptake and compartmentalization are similar in compatible pollinations on wild-type materials and in three style-side defective backgrounds: 120K-RNAi plants, HT-B-antisense plants (Supplementary Fig. S2a and b, respectively; Fig. 2c) and 4936-factor-mutant plants (Fig. 2f). Compatibility in these backgrounds is clearly associated with persistent sequestration of S-RNase. Furthermore, contrary to our earlier speculations about the possible functions of such non-S-RNase factors²⁰, 120K, HT-B and 4936-factor are not

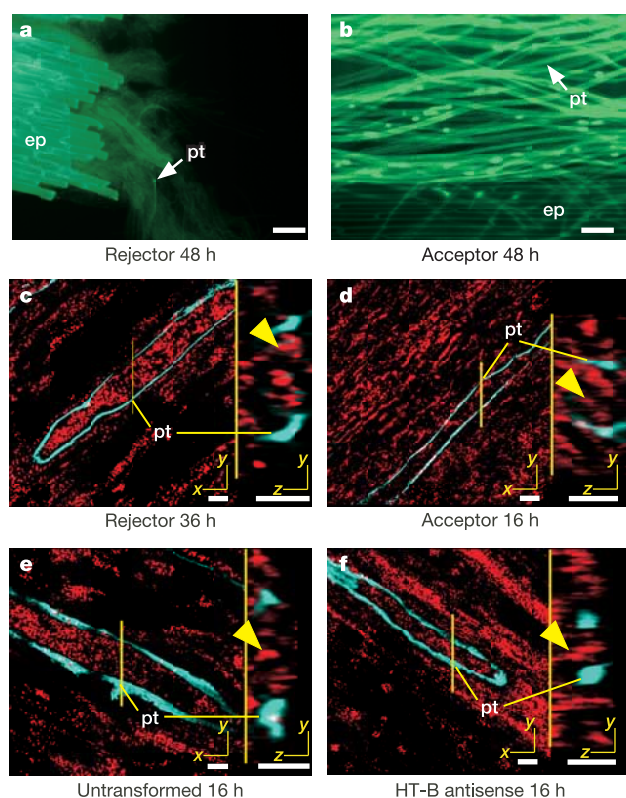


Figure 1 | S-RNase uptake in style-side defective backgrounds.

a, b, S_{105^-} -pollen tubes at the base of $S_{C10}S_{105^-}$ rejector and acceptor styles 48 h post-pollination. **c–f**, S-RNase uptake. Callose (cyan), S-RNase (red). Images illustrate the range of S-RNase labelling inside pollen tubes (**c**, heavy; **e, f**, intermediate; **d**, light). Left, single 300-nm optical sections in the x–y plane. Right, enlarged orthogonal projections viewed in the y–z plane. Vertical yellow lines show corresponding segments on the y-axis. Pollen tubes are indicated across both planes. Arrowheads, S-RNase inside pollen tubes (**c, e, f**, S_{C10^-} -RNase; **d**, S_{105^-} -RNase). Scale bars: **a, b**, 25 µm; **c–f**, 4 µm. ep, epidermis; pt, pollen tube.

required for S-RNase uptake. Instead, they seem to contribute to events that occur later, after S-RNase is inside the pollen tube.

The S-RNase compartment seems to break down during pollen rejection. As pollen tubes progress through the style, increasing numbers of incompatible tubes become distorted and show irregular cell walls and slow growth. Figure 2d shows an early-stage incompatible pollen tube 16 h post-pollination that does not show the distorted morphology characteristic of rejection; S-RNase compartmentalization is evident. Figure 2e shows a late-stage incompatible pollen tube at 36 h; the 120K signal is greatly reduced and there is no evidence for S-RNase compartmentalization. Labelling late-stage incompatible pollen tubes with vPPase antibody shows similar results (Supplementary Fig. S4). 120K and vPPase labelling is abundant in the surrounding transmitting tract, indicating that compartment breakdown specifically occurs in incompatible pollen tubes (Fig. 2e, Supplementary Fig. S4). Supplementary Movies S5 and S6 highlight the difference in S-RNase compartmentalization with three-dimensional animations of pollen tubes labelled with callose, S-RNase and 120K antibodies. Compartment breakdown was verified by examining 282 compatible or incompatible pollen tubes labelled with the 120K, callose and S_{C10} -RNase antibodies for intact versus disrupted compartments. Both normal and distorted pollen tubes (N and D, respectively; Fig. 2i) are usually present after

either compatible or incompatible pollination. The vast majority of compatible pollen tubes with normal morphology have an intact compartment. The proportion of distorted pollen tubes increases dramatically later in incompatible pollinations; these almost always show little or no evidence of compartmentalization. We also examined 66 pollen tubes labelled with vPPase, callose and S_{C10} -RNase antibodies 16 or 36 h after pollination. All 25 of the compatible pollen tubes showed normal morphology and an intact vPPase-labelled compartment. We examined 22 incompatible pollen tubes that did not yet show a distorted morphology, and all 22 also had intact compartments. Nineteen incompatible pollen tubes with distorted morphology were examined, but none of these showed an intact vPPase-labelled compartment. Our interpretation is that S-RNase is sequestered early in both compatible and incompatible pollination, and that the compartment is disrupted late in incompatible pollination.

S-RNase stability and specific degradation of HT-B

The absence, or near absence, of S-RNase from the compatible pollen tube cytoplasm could be explained either by a very limited release of S-RNase, or by degradation of non-self S-RNase. One study has reported lower S-RNase levels in compatible versus incompatible pollination in *Antirrhinum* and concluded that non-self S-RNase

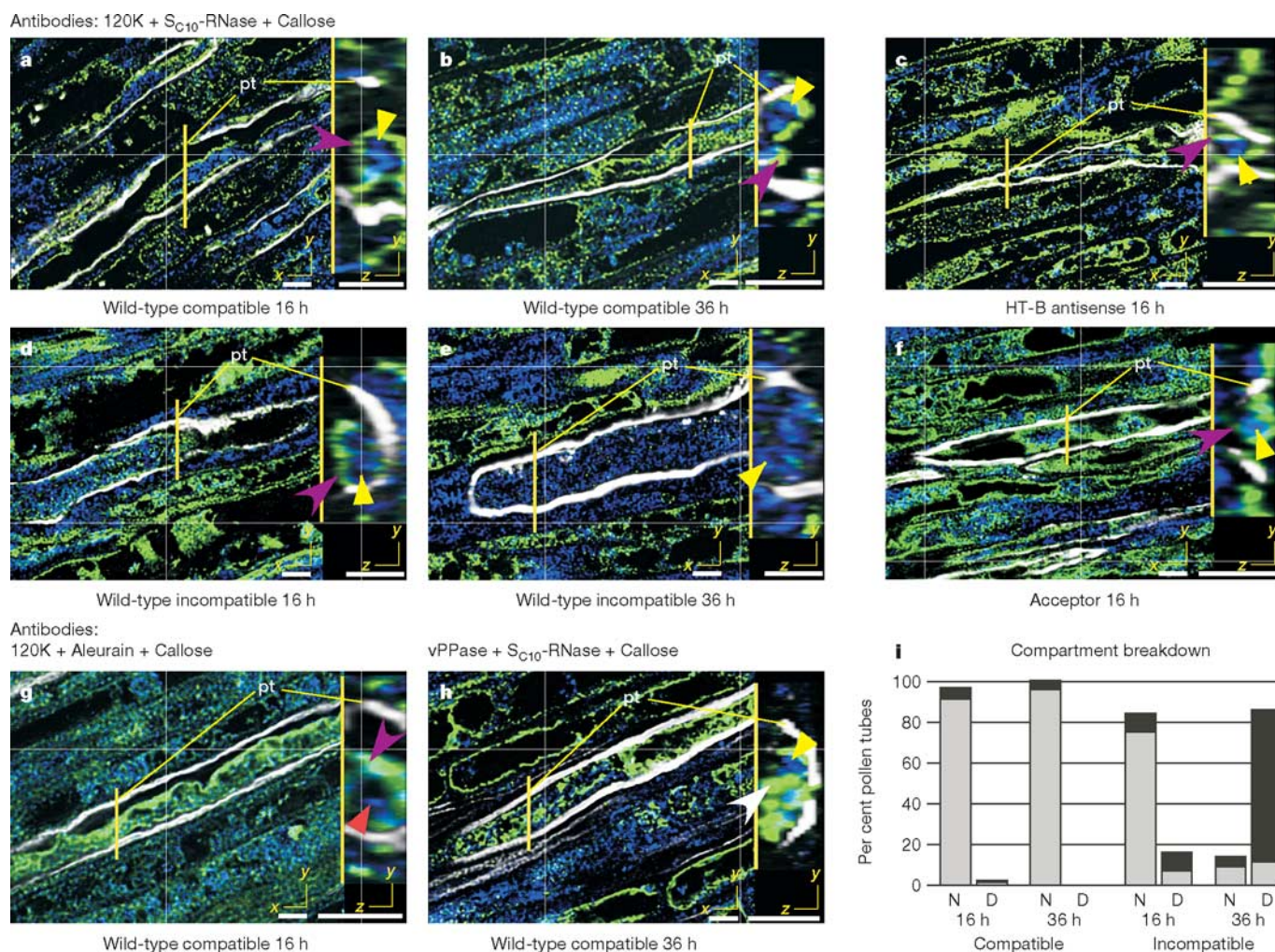


Figure 2 | S-RNase compartmentalization. a–h, 120K (green, purple arrow), callose (white), S_{C10} -RNase (blue, yellow arrow), aleurain (blue, red arrow), vPPase (green, white arrow). Orthogonal views are described in Fig. 1. The strong callose signal obscures 120K associated with the pollen wall. i, Normal (N) and distorted (D) pollen tubes showing intact

compartments. N and D sum to 100%. Shading shows percentage of N or D tubes with intact compartments: black, –compartment; grey, +compartment. For 16 h compatible, $n = 78$; 36 h compatible, $n = 79$; 16 h incompatible, $n = 68$; 36 h incompatible, $n = 57$. Scale bars, 4 μ m.

is degraded in compatible pollinations²⁹. We performed similar analyses and found that the average S-RNase level 36–48 h after compatible pollination is slightly lower than after incompatible pollinations, but the differences are well within the standard error and, in both cases, levels are higher than at time zero (Supplementary Fig. S7). Our interpretation is that the majority of S-RNase in pollen tubes is stable, but degradation of a minor proportion of S-RNase cannot be excluded.

Figure 3 shows that HT-B degradation in compatible pollen tubes is more prominent than is S-RNase degradation. $S_{A2}S_{A2}$ or $S_{105}S_{105}$ pistils were self-pollinated or pollinated with one of three compatible S-genotypes (Fig. 3a and b, respectively). Total pistil extracts were prepared, and both S-RNase and HT-B levels were measured to provide a basis for comparison. S-RNase levels were unchanged, but HT-B levels decreased about twofold after incompatible pollination. After compatible pollinations, the HT-B levels decreased 75–97%, compared to the control. Similar results were obtained with dissected pollen tubes (Supplementary Fig. S3b).

We used immunolocalization to investigate HT-B uptake and degradation in pollen tubes. The HT-B signals were much weaker in compatible compared to incompatible pollen tubes (Fig. 4a versus b). Figure 4c shows the results from scoring HT-B levels in 488 pollen tubes. Pollen tubes were given a score of 0 if little or no HT-B was present, and a score of 1, 2 or 3 for increasing amounts of HT-B (Supplementary Fig. S8). The inverse relationship is apparent—90% of compatible pollen tubes were scored as 0 or 1, while 83% of

incompatible pollen tubes were scored as 2 or 3 (Fig. 4c, grey and black bars, respectively). As HT-B is not S-specific, it is unlikely that this difference is due to selective uptake. Also, HT-B is not synthesized in pollen. Taken together, the immunolocalization and immunoblot results suggest that the decrease in steady-state HT-B levels is best explained by large-scale degradation of HT-B inside compatible pollen tubes. S-specific pollen rejection fails in HT-B-antisense plants with undetectable HT-B levels^{14,15}, but partially suppressed plants still show pollen rejection¹⁴. Therefore, large-scale degradation of HT-B by pollen tubes may allow them to escape rejection.

An alternative model for S-RNase-based SI

Figure 5 shows a model for compatible and incompatible pollinations (respectively, S_x - or S_a -pollen on an S_aS_b pistil) that incorporates these and previous results. S-RNase, 120K and HT-B are components of a selective pollen rejection system that pollen must overcome (by sequestering S-RNase and degrading HT-B) if it is to gain access to the ovules. Incompatibility is regarded as failure of the pollen's 'self-preservation' mechanisms: HT-B is stable and the S-RNase compartment breaks down, allowing S-RNase to exert its cytotoxic effect and resulting in pollen rejection. If, like most F-box proteins, SLF resides in the cytoplasm, then some S-RNase must be released for the proteins to interact, but this need not occur directly from the vacuole (arrows, Fig. 5). The binary cytotoxin ricin may provide a model for this release. Most ricin is sorted to the vacuole after entering the cell by endocytosis. About 5% moves through the Golgi and endoplasmic reticulum by retrograde transport. Less than 1% enters the cytoplasm^{35,36}.

Figure 5 shows SLF making both S-specific and non-S-specific interactions with S-RNase. Only non-S-specific interactions have been demonstrated biochemically²⁹, but the genetics of SI require S-specific interactions as well. In the current model, the non-self interaction leads to S-RNase degradation and pollen acceptance, and, conversely, the self interaction results in pollen rejection because S-RNase is not degraded^{19,11,29}. Instead, we propose that the self interaction leads to stabilization of HT-B, compartment breakdown and release of S-RNase. This alternative model leaves many

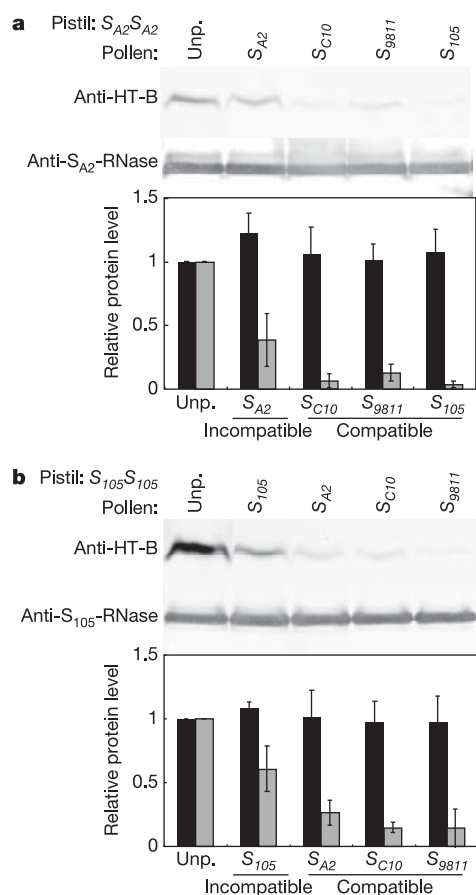


Figure 3 | Differential protein stability after compatible versus incompatible pollination. Specific degradation of HT-B in compatible pollinations in $S_{A2}S_{A2}$ (a) and $S_{105}S_{105}$ (b) pistils. Pistil extracts (total protein) were analysed to determine changes in HT-B (grey) and S-RNase (black) levels. Signals were normalized to unpollinated controls (Unp.). Sample immunoblots are shown for each S-genotype. Error bars indicate s.e.m.

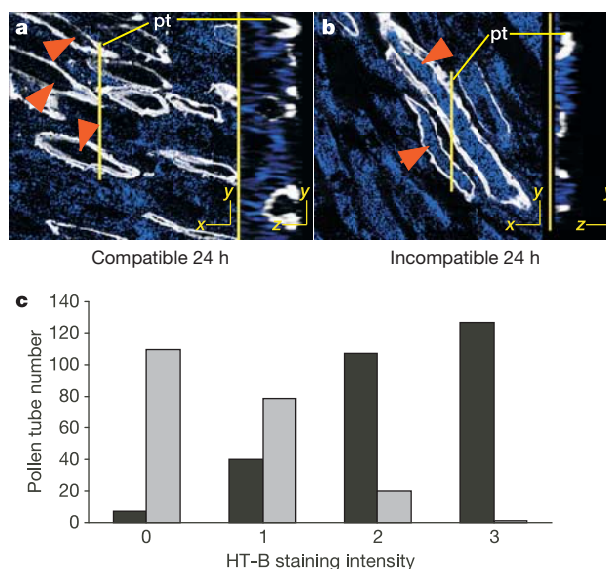


Figure 4 | HT-B in pollen tubes. a, b, Orthogonal views are described in Fig. 1. Callose (white), HT-B (blue). The HT-B signal is stronger in incompatible (arrows, b) than in compatible pollen tubes (arrows, a). c, HT-B immunolabelling in 280 incompatible (black) and 208 compatible (grey) pollen tubes was classified from low to high and designated 0, 1, 2 or 3, as shown in Supplementary Fig. S8.

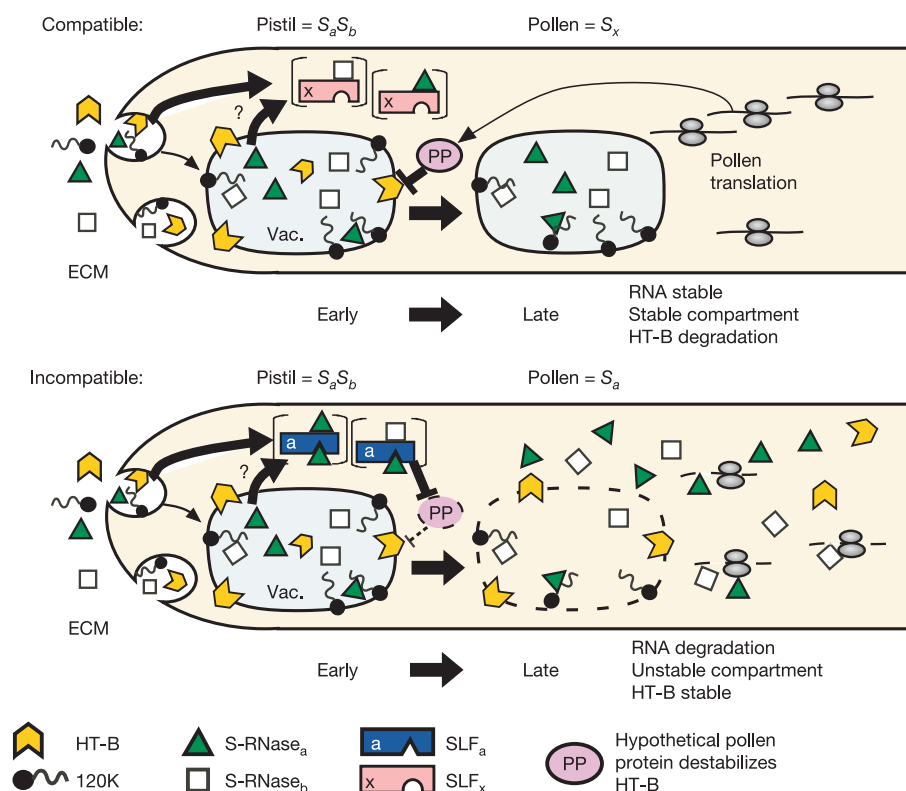


Figure 5 | A model for S-RNase-based SI. S-RNases, 120K and HT-B are taken up from the ECM by endocytosis. In compatible pollinations (top), pollen overcomes rejection by degrading HT-B and compartmentalizing S-RNase (Vac.). HT-B degradation is shown by a hypothetical pollen protein (PP), but other mechanisms are possible. Some S-RNase escapes and interacts with SLF (arrows, ?). In incompatible pollinations (bottom), the

self interaction ($S\text{-RNase}_a\text{-SLF}_a$) stabilizes HT-B, perhaps by inhibiting a factor such as PP. In the late stages of inhibition, the compartment breaks down, releasing S-RNase. RNA degradation prevents replacement of factors needed for degradation of HT-B and maintenance of the compartment. Thus, pollen rejection is self-reinforcing.

unanswered questions. It is not clear how the S-RNase/SLF interaction controls HT-B degradation and membrane breakdown. Figure 5 shows the self-S-RNase-SLF interaction stabilizing HT-B and preventing its degradation by a hypothetical pollen protein (PP; Fig. 5, right), but this is only one possibility. The effect could easily be indirect. The breakdown of the S-RNase compartment we observe may be similar to the cytoplasmic disorganization previously described in SI *Brugmansia suaveolens*³⁷. A role for S-like RNases in membrane stability has been proposed previously³⁸. However, our results show that the S-RNase-filled compartment is stable in the absence of stylar factors, such as HT-B. Thus, the events leading to breakdown of the compartment will require further study.

Our model recognizes three events contributing to S-specific pollen recognition and rejection: S-RNase uptake and sequestration, an S-specific recognition step that affects HT-B stability and, in incompatible pollen, a self-reinforcing cytotoxic step that results in rejection. Although only the second step is S-specific, each step involves interactions between pollen and pistil factors. S-RNase uptake is non-specific, and, in our model, both S-RNases in an S-heterozygote contribute to rejection once the internal compartment breaks down. In an incompatible cross, S-RNase cytotoxicity causes RNA degradation and interferes with pollen homeostasis, including functions such as S-RNase sequestration and HT-B degradation.

METHODS

Plant materials. SI *N. alata* ($S_{A2}S_{A2}$, $S_{C10}S_{C10}$, $S_{105}S_{105}$ and $S_{981}S_{981}$), as well as the HT-B antisense and controls, have been described previously^{14,39,40}. Generation of a population of plants segregating for the defect in 4936-factor is described in the Supplementary Notes.

Pollinations and stability studies. Flowers were emasculated before anthesis.

Style squashes in Fig. 1 were performed 48 h post-pollination, as described⁴¹. Pistil extracts were prepared for SDS-PAGE and immunostaining, also as described³³. Compatible crosses in Fig. 2 were ($S_{C10}S_{C10} \times S_{105}S_{105}$); incompatible crosses were ($S_{C10}S_{C10} \times S_{C10}S_{C10}$). In the experiment shown in Fig. 3, pistils were harvested for immunoblot analysis 72 h post-pollination. Heavy pollination was extremely important for analysis of HT-B stability in total pistil extracts. Less pollen was used when preparing material for HT-B immunolocalization (Fig. 4a $S_{105}S_{105} \times S_{C10}S_{C10}$; Fig. 4b $S_{C10}S_{C10} \times S_{C10}S_{C10}$). For the HT-B pollen tube counts in Fig. 4c, the compatible pollinations were $S_{A2}S_{A2} \times S_{C10}S_{C10}$, $S_{A2}S_{A2} \times S_{105}S_{105}$, $S_{105}S_{105} \times S_{C10}S_{C10}$ or $S_{C10}S_{C10} \times S_{105}S_{105}$; the incompatible pollinations were $S_{C10}S_{C10} \times S_{C10}S_{C10}$ or $S_{105}S_{105} \times S_{105}S_{105}$. **Immunolocalization.** The S_{C10} -RNase³⁹, aleurain³³ and 120K²⁰ antibodies have been described. Anti-callose (β -1,3-glucan) antibody was obtained from Biosupplies Australia. Anti- S_{105} -RNase antibody was prepared to MAP(8)-conjugated SQIEGAVDAVTKQL (Genemed Synthesis); anti-HT-B antibody was prepared to ovalbumin-conjugated DMVDPSISLLEPNNDKKTNG; anti-vPPase antibody was prepared to KLH-conjugated CLVGKVERNIPEDDPRN (Bethyl Labs). All antibodies were affinity purified. Pollinations were timed so that as many pollen tubes as possible would be present in the imaged region from 0.5–2.5 cm below the stigma. This region consistently provided the best tissue sections. Faster growing compatible pollen tubes were best visualized 16–24 h after pollination; more incompatible pollen tubes enter this region after 36 h. Double- and triple-labelling on fixed and paraffin-sectioned styles was performed, as described^{42,43}. When two primary antibodies from the same species were used, they were either directly labelled with Alexa 568 or Alexa 647 (Molecular Probes), or were used sequentially with an intermediate fixation step⁴². Otherwise, Alexa 488-, 568- or 647-conjugated secondary antibodies were used (Molecular Probes). Control sections lacking primary or secondary antibodies showed no signal. Additional details are provided in the Supplementary Information.

Received 9 July; accepted 23 November 2005.

- de Nettancourt, D. *Incompatibility and Incongruity in Wild and Cultivated Plants* (Springer, Berlin, 2001).

2. Kao, T.-h. & Tsukamoto, T. The molecular and genetic bases of S-RNase-based self-incompatibility. *Plant Cell* **16**, S72–S83 (2004).
3. Murfett, J., Atherton, T. L., Mou, B., Gasser, C. S. & McClure, B. A. S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection. *Nature* **367**, 563–566 (1994).
4. Lee, H.-S., Huang, S. & Kao, T.-h. S proteins control rejection of incompatible pollen in *Petunia inflata*. *Nature* **367**, 560–563 (1994).
5. Cornish, E. C., Pettitt, J. M., Bonig, I. & Clarke, A. E. Developmentally controlled expression of a gene associated with self-incompatibility in *Nicotiana alata*. *Nature* **326**, 99–102 (1987).
6. Anderson, M. A. *et al.* Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. *Plant Cell* **1**, 483–491 (1989).
7. McClure, B. S-RNase and SLF determine S-haplotype-specific pollen recognition and rejection. *Plant Cell* **16**, 2840–2847 (2004).
8. Entani, T. *et al.* Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: Identification of a pollen-expressed F-box gene with allelic diversity. *Genes Cells* **8**, 203–213 (2003).
9. Sijacic, P. *et al.* Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429**, 302–305 (2004).
10. Ikeda, K. *et al.* Primary structural features of the S haplotype-specific F-box protein, SFB, in *Prunus*. *Sex. Plant Reprod.* **16**, 235–243 (2004).
11. Ushijima, K. *et al.* The S haplotype-specific F-box protein gene, *SFB*, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. *Plant J.* **39**, 573–586 (2004).
12. Ushijima, K. *et al.* Structural and transcriptional analysis of the self-incompatibility locus of almond: Identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell* **15**, 771–781 (2003).
13. McClure, B. A., Cruz-Garcia, F., Beecher, B. S. & Sulaman, W. Factors affecting inter- and intra-specific pollen rejection in *Nicotiana*. *Ann. Bot.* **85**, 113–123 (2000).
14. McClure, B. A., Mou, B., Canevascini, S. & Bernatzky, R. A small asparagine-rich protein required for S-allele-specific pollen rejection in *Nicotiana*. *Proc. Natl Acad. Sci. USA* **96**, 13548–13553 (1999).
15. O'Brien, M. *et al.* Molecular analysis of the stylar-expressed *Solanum chacoense* asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. *Plant J.* **32**, 1–12 (2002).
16. Kondo, K. *et al.* Cultivated tomato has defects in both S-RNase and HT genes required for stylar function of self-incompatibility. *Plant J.* **29**, 627–636 (2002).
17. Kondo, K. *et al.* Insights into the evolution of self-compatibility in *Lycopersicon* from a study of stylar factors. *Plant J.* **30**, 143–153 (2002).
18. Lind, J. L., Bacic, A., Clarke, A. E. & Anderson, M. A. A style-specific hydroxyproline-rich glycoprotein with properties of both extensins and arabinogalactan proteins. *Plant J.* **6**, 491–502 (1994).
19. Lind, J. L., Bönig, I., Clarke, A. E. & Anderson, M. A. A style-specific 120 kDa glycoprotein enters pollen tubes of *Nicotiana alata* in vivo. *Sex. Plant Reprod.* **9**, 75–86 (1996).
20. Cruz-Garcia, F., Hancock, C. N., Kim, D. & McClure, B. Stylar glycoproteins bind to S-RNase in vitro. *Plant J.* **42**, 295–304 (2005).
21. Hancock, C. N., Kent, L. & McClure, B. The 120 kDa glycoprotein is required for S-specific pollen rejection in *Nicotiana*. *Plant J.* **43**, 716–723 (2005).
22. Martin, F. W. The behaviour of *Lycopersicon* incompatibility alleles in an alien genetic milieu. *Genetics* **60**, 101–109 (1968).
23. Ai, Y., Kron, E. & Kao, T.-h. S-alleles are retained and expressed in a self-compatible cultivar of *Petunia hybrida*. *Mol. Gen. Genet.* **230**, 353–358 (1991).
24. Tsukamoto, T. *et al.* Breakdown of self-incompatibility in a natural population of *Petunia axillaris* (Solanaceae) in Uruguay containing both self-incompatible and self-compatible plants. *Sex. Plant Reprod.* **12**, 6–13 (1999).
25. Tsukamoto, T. *et al.* Breakdown of self-incompatibility in a natural population of *Petunia axillaris* caused by a modifier locus that suppresses the expression of an S-RNase gene. *Sex. Plant Reprod.* **15**, 255–263 (2003).
26. Bernatzky, R., Glaven, R. H. & Rivers, B. A. S-related protein can be recombined with self-compatibility in interspecific derivatives of *Lycopersicon*. *Biochem. Genet.* **33**, 215–225 (1995).
27. Luu, D.-T., Xike, Q., Morse, D. & Cappadocia, M. S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature* **407**, 649–651 (2000).
28. McClure, B. A., Gray, J. E., Anderson, M. A. & Clarke, A. E. Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. *Nature* **347**, 757–760 (1990).
29. Qiao, H. *et al.* The F-box protein AhSLF-S₂ physically interacts with S-RNases that may be inhibited by the ubiquitin/26S proteasome pathway of protein degradation during compatible pollination in *Antirrhinum*. *Plant Cell* **16**, 582–595 (2004).
30. Sonneveld, T., Tobutt, K. R., Vaughan, S. P. & Robbins, T. P. Loss of pollen-S function in two self-compatible selections of *Prunus avium* is associated with deletion/mutation of an S haplotype-specific F-box gene. *Plant Cell* **17**, 37–51 (2005).
31. Qiao, H. *et al.* The F-box protein AhSLF-S₂ controls the pollen function of S-RNase-based self-incompatibility. *Plant Cell* **16**, 2307–2322 (2004).
32. Hicks, G. R., Rojo, E., Hong, S., Carter, D. C. & Raikhel, N. V. Germinating pollen has tubular vacuoles, displays highly dynamic vacuole biogenesis, and requires VACUOLESS1 for proper function. *Plant Physiol.* **134**, 1227–1230 (2004).
33. Rogers, S. W., Burks, M. & Rogers, J. C. Monoclonal antibodies to barley aleurain and homologs from other plants. *Plant J.* **11**, 1359–1368 (1997).
34. Jauh, G.-Y., Phillips, T. E. & Rogers, J. C. Tonoplast intrinsic protein isoforms as markers for vacuolar functions. *Plant Cell* **11**, 1867–1882 (1999).
35. van Deurs, B. *et al.* Estimation of the amount of internalized ricin that reaches the trans-golgi network. *J. Cell Biol.* **106**, 253–267 (1988).
36. Wesche, J., Rapak, A. & Olsnes, S. Dependence of ricin toxicity on translocation of the toxin A-chain from the endoplasmic reticulum to the cytosol. *Proc. Natl Acad. Sci. USA* **274**, 34443–34449 (1999).
37. Geitmann, A. in *Fertilization in Higher Plants* (eds Cresti, M., Cai, G. & Moscatelli, A.) (Springer, Berlin/Heidelberg, 1999).
38. Macintosh, G. C., Bariola, P. A., Newbiggin, E. & Green, P. J. Characterization of Rny1, the *Saccharomyces cerevisiae* member of the T₂ RNase family of RNases: Unexpected functions for ancient enzymes? *Proc. Natl Acad. Sci. USA* **98**, 1018–1023 (2001).
39. Murfett, J. M. *et al.* S-RNase and interspecific pollen rejection in the genus *Nicotiana*: Multiple pollen rejection pathways contribute to unilateral incompatibility between self-incompatible and self-compatible species. *Plant Cell* **8**, 943–958 (1996).
40. Beecher, B. & McClure, B. A. Effects of RNases on rejection of pollen from *Nicotiana tabacum* and *N. plumbaginifolia*. *Sex. Plant Reprod.* **14**, 69–76 (2001).
41. Kho, Y. O. & Baer, J. Observing pollen tubes by means of fluorescence. *Euphytica* **17**, 299–302 (1968).
42. Paris, N. & Rogers, J. C. The role of receptors in targeting soluble proteins from the secretory pathway to the vacuole. *Plant Physiol. Biochem.* **34**, 223–237 (1996).
43. Tanner, V. A., Ploug, T. & Tao-Cheng, J. H. Subcellular localization of SV2 and other secretory vesicle components in PC12 cells by an efficient method of preembedding EM immunocytochemistry for cell cultures. *J. Histochem. Cytochem.* **44**, 1481–1488 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Kroll for manuscript assistance; H. Sassa for helpful discussions; J. Rogers for antibodies; and C. S. Gasser, S. J. Hiscock and C. Staiger for pre-review. The MU-Monsanto Plant Biology Program and grants from the US National Science Foundation supported this work. F.C.-G. was also supported by a grant from the Universidad Nacional Autonoma de Mexico.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.M. (mcclureb@missouri.edu).

Histone demethylation by a family of JmjC domain-containing proteins

Yu-ichi Tsukada^{1,2}, Jia Fang^{1,2}, Hediye Erdjument-Bromage³, Maria E. Warren², Christoph H. Borchers², Paul Tempst³ & Yi Zhang^{1,2}

Covalent modification of histones has an important role in regulating chromatin dynamics and transcription. Whereas most covalent histone modifications are reversible, until recently it was unknown whether methyl groups could be actively removed from histones. Using a biochemical assay coupled with chromatography, we have purified a novel JmjC domain-containing protein, JHDM1 (JmjC domain-containing histone demethylase 1), that specifically demethylates histone H3 at lysine 36 (H3-K36). In the presence of Fe(II) and α -ketoglutarate, JHDM1 demethylates H3-methyl-K36 and generates formaldehyde and succinate. Overexpression of JHDM1 reduced the level of dimethyl-H3-K36 (H3K36me2) *in vivo*. The demethylase activity of the JmjC domain-containing proteins is conserved, as a JHDM1 homologue in *Saccharomyces cerevisiae* also has H3-K36 demethylase activity. Thus, we identify the JmjC domain as a novel demethylase signature motif and uncover a protein demethylation mechanism that is conserved from yeast to human.

Covalent histone modifications have an important role in regulating chromatin dynamics and function¹. One such modification, methylation, occurs on both lysine and arginine residues and participates in a diverse range of biological processes including heterochromatin formation, X-chromosome inactivation and transcriptional regulation^{2–4}. Unlike acetylation, which generally correlates with transcriptional activation, histone lysine methylation can signal either activation or repression depending on the particular lysine residue that is methylated⁴. Even within the same lysine residue, the biological consequence of methylation can differ depending on whether the lysine residue contains mono-, di-, or trimethylation^{5,6}.

The steady-state level of a covalent histone modification is controlled by a balance between enzymes that catalyse the addition and removal of a given modification. Although this notion is generally true for many histone modifications, an enzyme capable of removing methyl groups from a methyl-lysine residue has remained elusive, until recently⁷. Using a candidate approach, it has been demonstrated that LSD1/BHC110, a nuclear amine oxidase homologue previously found in several histone deacetylase complexes^{8–10}, can specifically demethylate monomethyl-H3-K4 (H3K4me1) and H3K4me2 in a FAD (flavin adenine dinucleotide)-dependent oxidative reaction. Although potential LSD1 homologues exist in *Schizosaccharomyces pombe*, they are not found in *S. cerevisiae*, even though at least three lysine residues on H3 can be methylated in this organism. This observation, and that the LSD1 family demethylases are unable to demethylate trimethyl-lysine residues, raises the possibility that additional demethylases that use a different reaction mechanism exist. Using a novel biochemical assay, we have identified a histone demethylase activity from HeLa cells. Here we describe the purification, identification and functional characterization of a family of H3-K36 demethylases. Our results indicate that the JmjC domain, a motif conserved from *S. cerevisiae* to human, is a signature motif for this family of histone demethylases.

Histone demethylase activity in HeLa cell extracts

Recent studies have demonstrated that the methyl groups of

1-methyladenine (1-meA) and 3-methylcytosine (3-meC) in DNA can be removed by the AlkB family of proteins through oxidative demethylation^{11,12}. The similarity between the chemistry of removing a methyl group from 1-meA and methyl-lysine (Fig. 1a) prompted us to test whether a similar mechanism is used in histone lysine demethylation. We developed an *in vitro* assay based on detection of formaldehyde, one of the predicted release products. To maximize detection sensitivity, radiolabelled nucleosomal histones were used as substrates. As outlined in Supplementary Fig. S1, labelled substrates were subjected to demethylation reactions in the presence of cofactors Fe(II) and α -ketoglutarate. Released [³H]-formaldehyde was converted to radioactive 3,5-diacetyl-1, 4-dihydrolutidine before being extracted and counted.

Using the assay described above, we analysed the protein fractions derived from HeLa cell nuclear extracts and nuclear pellet¹³. Results shown in Fig. 1b indicate that a demethylase activity is enriched in the nuclear pellet-derived 0.3 M P11 protein fraction. This demethylase activity is dependent on the presence of the protein fraction (Fig. 1c, lane 1), as well as the predicted cofactors (Fig. 1c, lanes 3 and 4). In addition, ascorbate is also required for optimal activity, probably due to its ability to regenerate Fe(II) from Fe(III). These results suggest the presence of a histone demethylase activity in the 0.3 M P11 fraction.

Histone demethylase activity of a JmjC domain-containing protein

To identify the protein(s) responsible for the demethylase activity, we monitored the enzymatic activity through six chromatography columns (Supplementary Fig. S2). After purification of the 0.3 M P11 fraction through DEAE5PW and Phenyl Sepharose columns, we determined the native mass of the enzymatic activity to be about 300 kDa on a Superose 6 column (Fig. 2a). Further purification on a MonoQ column allowed us to correlate two protein bands (marked by an asterisk in the top panel of Fig. 2b) with the enzymatic activity (Fig. 2b, bottom panel). Mass spectrometry analysis identified the larger of the two proteins as F-box and leucine-rich repeat protein 11

¹Howard Hughes Medical Institute, ²Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA. ³Molecular Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

(FBXL11) (Fig. 2c). The identity of the smaller band was inconclusive based on the established criteria for protein identification¹⁴.

FBXL11 is an uncharacterized protein that was originally identified through bioinformatic searches for F-box-containing proteins^{15,16}. In

addition to an F-box, FBXL11 contains several interesting domains including a JmjC domain, a CxxC zinc finger, a PHD domain and three leucine-rich repeats (Fig. 3a). JmjC domain-containing proteins are predicted to be metalloenzymes that regulate chromatin function¹⁷. This prediction, and that the demethylase activity requires Fe(II) as a cofactor, are consistent with FBXL11 being the responsible enzyme. We confirmed this possibility by demonstrating that Flag-tagged FBXL11 immunoprecipitated from transfected COS-7 cells exhibits robust histone demethylase activity (Fig. 3b, compare lanes 1 and 2).

To evaluate which domains of FBXL11 are important for its enzymatic activity, we generated a series of expression constructs with deletions of the JmjC domain, CxxC zinc finger, PHD domain, F-box, or leucine-rich repeat, respectively. After transfection and immunoprecipitation, these mutant proteins were subjected to western blot analysis (Fig. 3b, top panel) and demethylase activity assays (Fig. 3b, bottom panel). The results demonstrate that only the JmjC domain is absolutely required, although deletion of the CxxC zinc finger, PHD domain and the leucine-rich repeats also partially impaired the enzymatic activity (compare lanes 4–8 with 2). To demonstrate further the importance of the JmjC domain for enzymatic activity, we generated an H212A mutant. We choose to mutate H212 not only because this histidine is highly conserved in the JmjC domain of FBXL11 orthologues (Supplementary Fig. S3b), but also because the corresponding histidine in FIH (factor-inhibiting hypoxia-inducible factor), a known Fe(II)-dependent oxygenase, directly binds to Fe(II)¹⁸. Results shown in Fig. 3b indicate that the H212A mutation completely abolishes the enzymatic activity of FBXL11 (compare lanes 2 and 3). On the basis of the above results we conclude that FBXL11 is a novel histone demethylase and that the JmjC domain is critical for its enzymatic activity. Because histone demethylase activity is the first function attributed to FBXL11 and because FBXL11 is the first JmjC domain-containing protein shown to possess histone demethylase activity, we have named the protein JHDM1A (JmjC domain-containing histone demethylase 1A.). The highly related protein FBXL10, which we have named JHDM1B (Supplementary Fig. S3a), is also an active H3-K36 demethylase (data not shown).

JHDM1A preferentially demethylates H3K36me2

To characterize further JHDM1A, we generated and purified recombinant Flag-JHDM1A from baculovirus-infected Sf9 cells (Supplementary Fig. S4). We analysed its site specificity using histone substrates radiolabelled at all known methylated sites in histones H3 (K4, K9, K27, K36, K79) and H4 (K20, R3). Of the seven substrates, only H3-K36 methylated by Set2 was a substrate for JHDM1A (Fig. 3c). The substrate specificity was also verified by comparing the methylation state of nucleosomal histone substrate after subjecting to JHDM1A-mediated demethylation (Supplementary Fig. S5). Thus, we conclude that JHDM1A is an H3-K36-specific demethylase.

Lysine residues exist in three methylation states (mono-, di- and trimethylation). To determine whether JHDM1A preferentially demethylates a particular methylation state, free histones, mononucleosomes and oligonucleosomes were subjected to a demethylation reaction in the presence or absence of recombinant JHDM1A. Results shown in Fig. 3d indicate that JHDM1A preferentially demethylates K36me2 (second panel, compare lanes 2, 4 and 6 with 1, 3 and 5), although a decrease in K36me1 levels was also observed (first panel). In contrast, no change in trimethyl-K36 (K36me3) levels was detected. Under these assay conditions, JHDM1A was capable of demethylating H3-K36 regardless of whether it was in free, mono-, or oligonucleosome form. To analyse further the K36me2 selectivity of JHDM1A, K36me2 or K36me3 peptides were subjected to demethylation reactions. Mass spectrometry analysis revealed the presence of mono- and unmethylated forms of the peptide in a JHDM1A-dependent manner (Fig. 3e). Demethylation of the peptide occurs in a time- and concentration-dependent manner

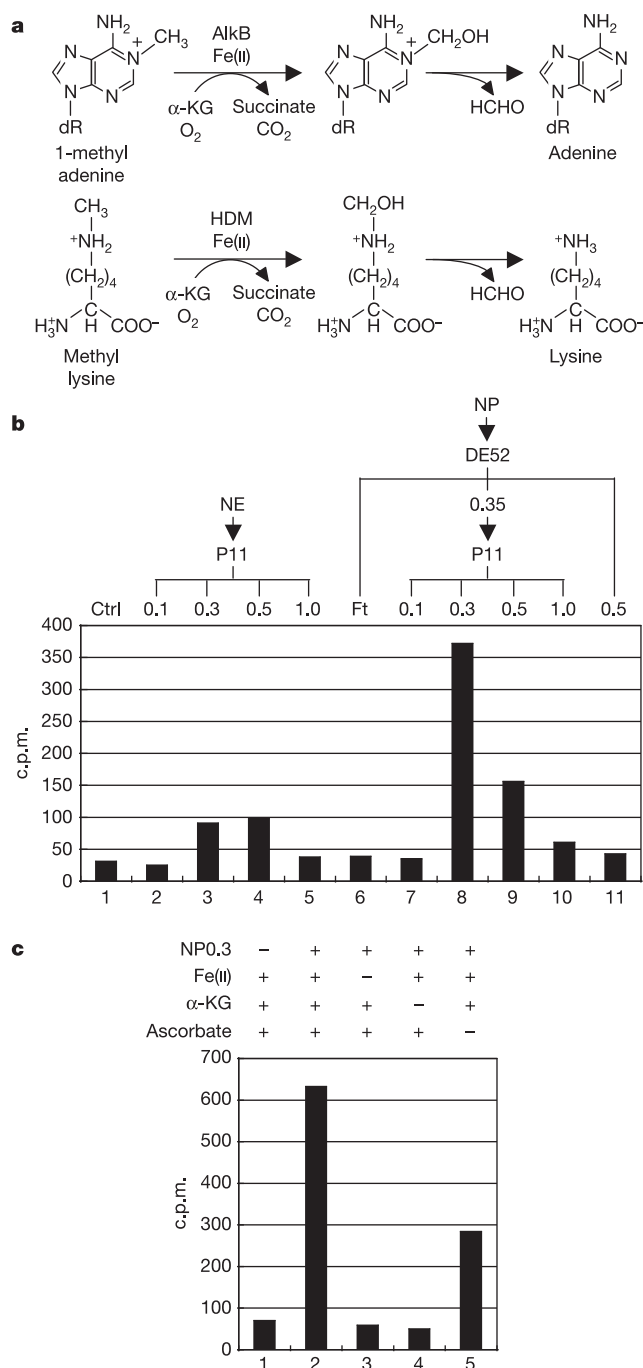


Figure 1 | Identification of a demethylase activity in HeLa cells.

a, Schematic presentation of the demethylation mechanism of 1-methyladenine by the AlkB family of proteins (top panel). The bottom panel shows the proposed mechanism for histone demethylation. For simplicity, only monomethyl-lysine is illustrated, but the same mechanism can be applied to di-, or trimethylated lysine residues. α -KG, α -ketoglutarate. **b**, Histone demethylase activities in the P11 column fractions derived from HeLa nuclear extracts (NE) and nuclear pellet (NP) fractions. The substrates used are Set2-methylated nucleosomes. The numbers above the panel represent the molar concentration of KCl in the elution buffers. c.p.m., counts per minute. Ft, flow through fraction. **c**, The demethylase activity depends on the presence of Fe(II), α -ketoglutarate and proteins present in the 0.3 M P11 nuclear pellet fraction.

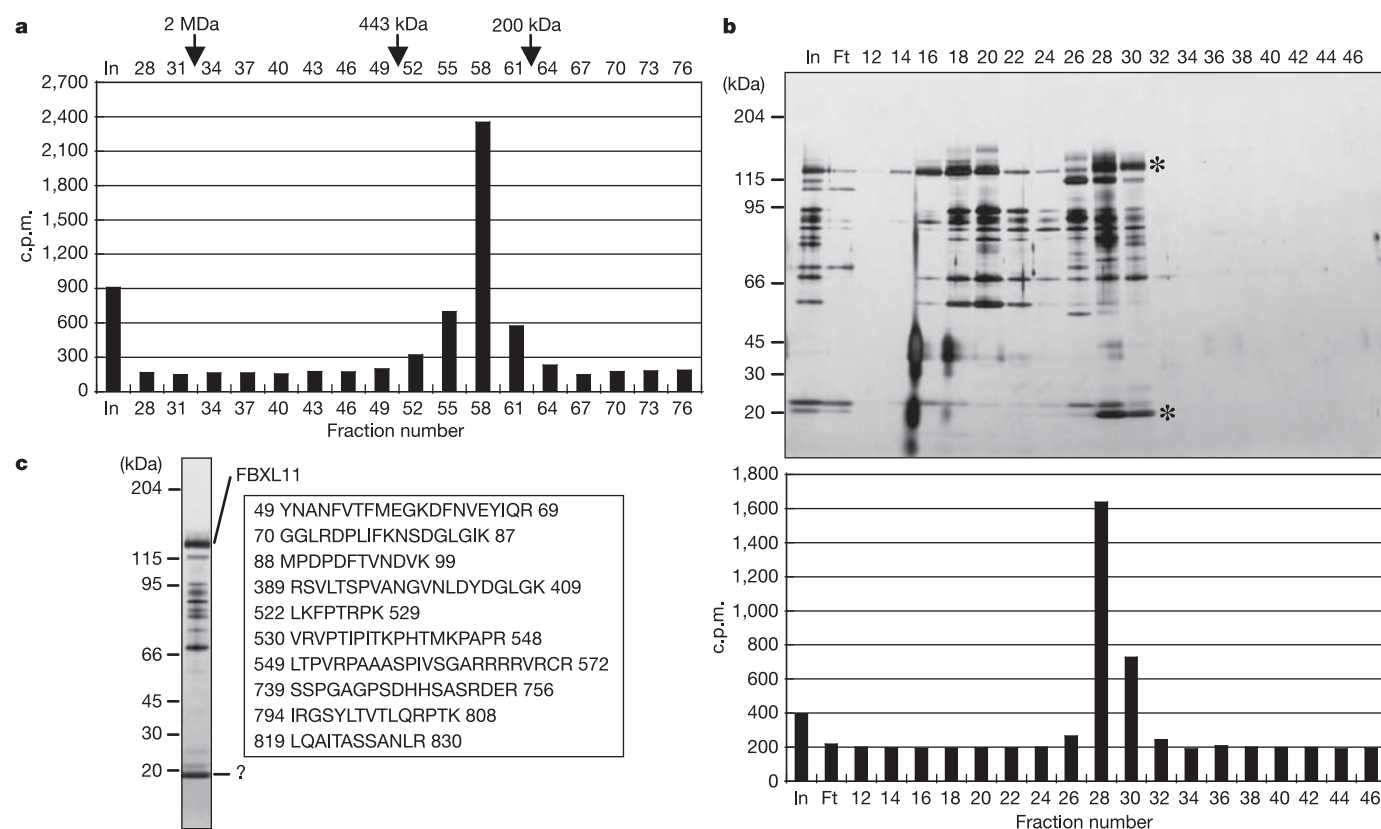


Figure 2 | Purification and identification of a histone demethylase activity. **a**, Histone demethylase activities of the protein fractions derived from a gel-filtration Superose 6 column. The elution profile of the protein markers is indicated along the top of the panel. **b**, A silver-stained protein gel (top panel) and histone demethylase activities (bottom panel) of the protein fractions derived from a mini MonoQ column. The candidate proteins that co-fractionated with the demethylase activity are indicated by an asterisk.

The positions of the protein size markers on SDS-PAGE are indicated to the left of the panel. **c**, A silver-stained protein gel containing the samples for protein identification. A total of 34 tryptic peptides covering 31% of FBXL11 (NP_036440) were identified by mass fingerprinting for the top band, and 10 representative peptides identified are listed. Numbers correspond to the amino acid location within the FBXL11 protein. The question mark represents an unidentified FBXL11-associated protein.

(Supplementary Fig. S6). In contrast, no demethylation was detected when a K36me3 peptide was subjected to a parallel reaction (data not shown). Collectively, the above results indicate that JHDM1A selectively demethylates H3-K36 with a preference for the dimethyl form.

To study the effect of JHDM1A on H3-K36 methylation *in vivo*, we transfected a mammalian Flag-JHDM1A expression vector into 293T cells. Immunofluorescence staining revealed that overexpression of wild-type Flag-JHDM1A, but not the H212A mutant, results in a significant decrease in the level of K36me2 (Fig. 3f). However, overexpression of Flag-JHDM1A does not alter the level of K36me1, K36me3, nor the level of K4me2 (Supplementary Fig. S7). These results indicate that JHDM1A demethylates H3K36me2 *in vivo*.

Demethylation by JHDM1A generates formaldehyde and succinate

The demethylation reaction is predicted to generate formaldehyde and succinate (Fig. 1a). We confirmed this prediction with mass spectrometry. Under the assay conditions, formaldehyde would exist in its protonated form with a mass-to-charge (m/z) ratio of 31.0184. Consistent with formaldehyde as a reaction product, an ion with an m/z ratio of 31.0239 was detected when the reaction was carried out in the presence, but not in the absence, of JHDM1A (Fig. 4a). Using a similar approach, we also detected an ion (m/z 119.0778) that correlated with protonated succinate ($C_4O_4H_7$) in a JHDM1A-dependent manner (Fig. 4b). Because an ion of similar m/z ratio—119.1215, although lower in abundance—was also detected in the reaction mixture in the absence of JHDM1A (Fig. 4b, bottom panel), we sought to verify that the ion detected in the presence of JHDM1A was indeed succinate. This was achieved by demonstrating that the

presumed succinate ion produced in the demethylation reaction generated the identical fragmentation MS/MS spectrum to pure succinate (Fig. 4c). In contrast, no ion was detected in the absence of JHDM1A (data not shown), indicating that the lower abundance ion detected in the control reaction (Fig. 4b, bottom panel) was not succinate. Given that Fe(II)- and α -ketoglutarate-dependent enzymes often catalyse a slow, uncoupled reaction in the absence of substrates, we analysed formation of succinate in the absence of a K36me2 peptide. As expected, JHDM1A catalyses low levels of uncoupled reaction in the absence of the substrate peptide, but production of succinate is stimulated by the presence of the peptide (Supplementary Fig. S8). These results support the hypothesis that JHDM1A-mediated histone demethylation generates formaldehyde and succinate, thus confirming the proposed demethylation mechanism.

Conserved JmjC domain-mediated histone demethylation

Given that JmjC domain-containing proteins exist in organisms from bacteria to human (<http://smart.embl-heidelberg.de>), we asked whether the demethylase activity of this domain is conserved. In addition to the highly related JHDM1B found in human and mouse, potential JHDM1A homologues were identified in many other eukaryotic organisms including *Xenopus*, *Drosophila*, *Caenorhabditis elegans*, *S. pombe* and *S. cerevisiae* (Supplementary Fig. S3a). Although the domain structures are not completely conserved among these proteins, their JmjC domains are highly conserved (Supplementary Fig. S3b). Importantly, alignment of their JmjC domains with that of FIH revealed strict conservation—with the exception of the *S. pombe* homologue—of the amino acids, marked

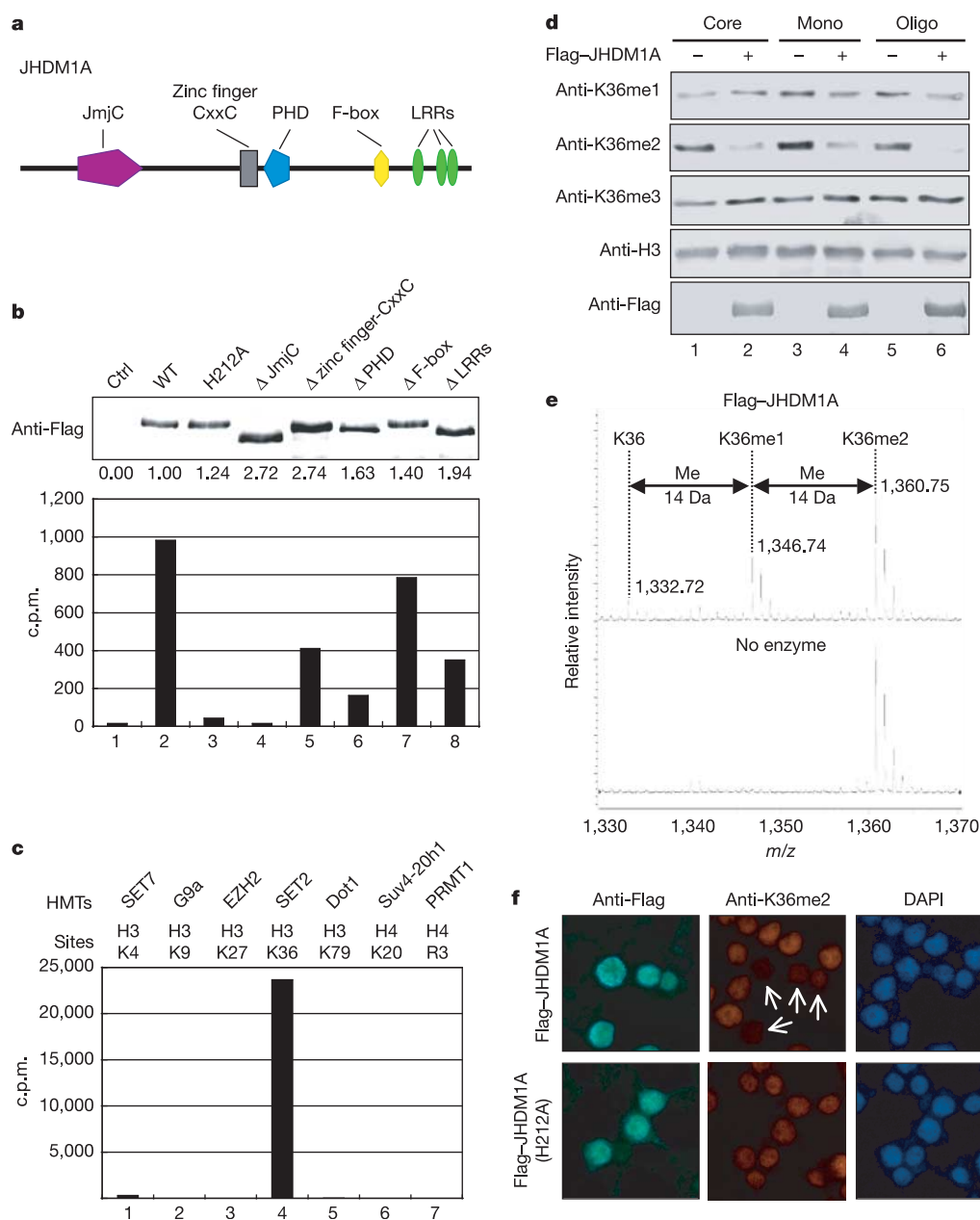


Figure 3 | Characterization of the histone demethylase JHDM1A/FBXL11. **a**, Schematic representation of the JHDM1A protein. Functional domains were identified using the SMART program (http://smart.embl-heidelberg.de). **b**, Wild-type and mutant Flag-JHDM1A proteins were expressed in COS-7 cells, and immunoprecipitated samples were divided and analysed by western blotting (top panel) and demethylase assay (bottom panel). The relative amount of each immunoprecipitated protein, indicated with numbers below the western blot, was quantified with ImageJ (http://rsb.info.nih.gov/ij) and used for normalization of the activities. **c**, Histone demethylase activity of recombinant Flag-JHDM1A, purified from baculovirus-infected Sf9 cells, against various methylated histone

substrates. The histone methyltransferases (HMTs) and their sites of methylation are indicated above the panel. The demethylase assays were performed with equal inputs, either by counts or by histone amounts, with similar results. **d**, Western blot analysis of demethylation reactions using various histone substrates. Antibodies used are indicated to the left of the panel. **e**, Mass spectrometry analysis of demethylation of a K36me2 peptide (STGGV2mKKPHRY-C) by purified Flag-JHDM1A. The enzyme/substrate molar ratio of the reaction is 1:40. Numbers represent the masses of the substrate and product peptides. **f**, Overexpression of wild-type (top panels) but not a demethylase-defective mutant (bottom panels) of Flag-JHDM1A in 293T cells resulted in decreased levels of H3K36me2 as marked by arrows.

by an asterisk and a hash symbol, that are involved in Fe(II)- and α -ketoglutarate-binding, respectively¹⁸, indicating that these proteins are likely to be active demethylases. To examine this possibility, we cloned JHDM1B and demonstrated its H3-K36 demethylase activity (data not shown). We also cloned and expressed YER051W, which encodes the *S. cerevisiae* homologue scJHDM1. Recombinant scJHDM1 specifically demethylates methyl-K36, but not methyl-K4 or methyl-K79 (Fig. 5a). The detected demethylase activity is dependent on the JmjC domain, because a point mutation in the JmjC domain (H305A) completely abolished its enzymatic activity

(Fig. 5b). These results confirm the conserved function of the JmjC domain in histone demethylation.

Of the JHDM1 family members identified (Supplementary Fig. S3a), only the *S. pombe* homologue Epe1 has been functionally characterized¹⁹. In the absence of Epe1, centromeres and mating type loci are destabilized, with concomitant changes in methylation patterns of H3-K4 and H3-K9 (ref. 19). Normal function of Epe1 requires an intact JmjC domain, as the JmjC domain mutant Y307A failed to complement a loss-of-function Epe1 mutant. Notably, a tyrosine corresponding to Y307 in Epe1 is conserved in all the

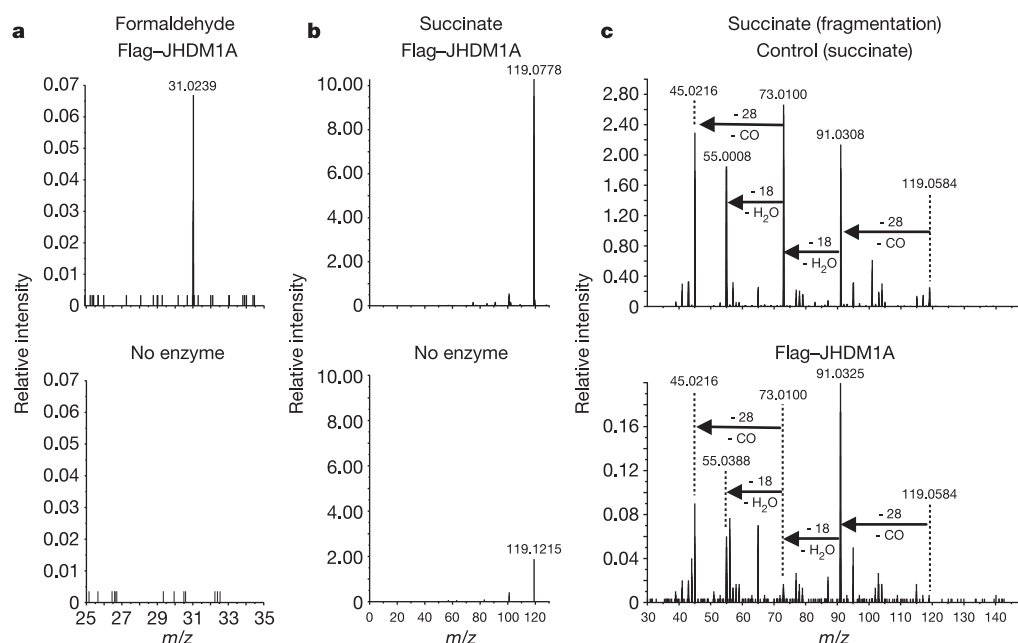


Figure 4 | Flag-JHDM1A-mediated histone demethylation generates formaldehyde and succinate. Demethylation reactions using highly purified Flag-JHDM1A expressed in Sf9 cells and a K36me2 peptide (STGGV2mKKPHRY-C) were subjected to mass spectrometric analysis. **a, b**, ESI-MS detection of formaldehyde (**a**; $[M + H]^+_{\text{theoretical}} = 31.0184$)

and succinate (**b**; $[M + H]^+_{\text{theoretical}} = 119.0344$) in the presence (upper panel) and absence (lower panel) of Flag-JHDM1A. **c**, ESI-MS/MS analysis of ions at m/z 119 from a standard solution of succinate (600 nM) (top panel), and a Flag-JHDM1A reaction sample (bottom panel). Suggested structures of the succinate fragment ions are shown on the MS/MS spectrum.

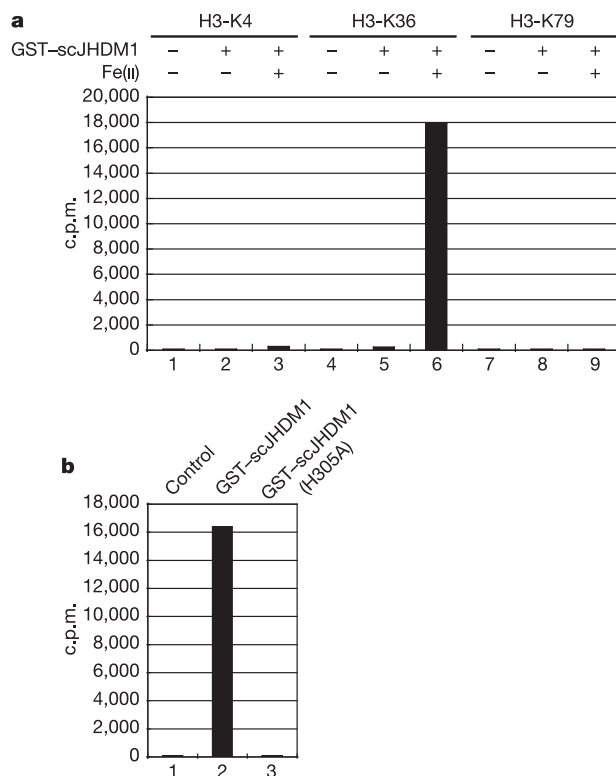


Figure 5 | The *S. cerevisiae* JHDM1 homologue exhibits H3-K36 demethylase activity. **a**, Demethylase activity and site specificity of the *S. cerevisiae* JHDM1 homologue expressed in *Escherichia coli*. **b**, Mutational analysis of the JmjC domain of the scJHDM1 protein. Equal amounts of wild-type and mutant GST-scJHDM1 were used in the demethylase assays. A mutation in the Fe(II)-binding site (H305A) completely abolished the H3-K36 demethylase activity.

JHDM1 family proteins (Supplementary Fig. S3b, marked by '\$'), suggesting involvement of this residue in a conserved function, such as histone demethylation. Surprisingly, Epe1 lacks conservation of two residues in the JmjC domain predicted to bind to Fe(II) (Supplementary Fig. S3b). This raises the question of whether Epe1 is an active histone demethylase. Results shown in Supplementary Fig. S9 indicate that recombinant Epe1 purified from Sf9 cells lacks histone demethylase activity towards any of the four lysine residues on *S. pombe* H3 known to be subjected to methylation. Whether native Epe1 is a histone demethylase remains to be determined, but our *in vitro* evidence suggests that the phenotypes associated with Epe1 mutation may not be directly related to histone demethylation. Instead, the phenotypes may result from other affects on histone methyltransferase recruitment or chromatin remodelling.

The JmjC domain is a signature motif for histone demethylases

By monitoring histone demethylase activity, we purified an H3-K36 demethylase and demonstrated that its JmjC domain is critical for its enzymatic activity. This result together with the fact that the JmjC domain is involved in binding to the reaction cofactors Fe(II) and α -ketoglutarate¹⁸, allow us to conclude that the JmjC domain represents a signature motif for a new family of histone demethylases. Our results are consistent with a structure-based prediction that JmjC domain-containing proteins may be histone demethylases²⁰. In addition, our study also indicates that histone demethylation catalysed by JmjC domain proteins uses the same oxidative demethylation mechanism used by the AlkB family of DNA demethylases^{11,12}. It is worth mentioning that recent studies have demonstrated that methyl-arginine is turned over by a peptidyl arginine deiminase PADI4/PAD4, which converts methyl-arginine to citrulline^{21,22}. Because this reaction involves a change in the amino acid residue, it is not a strict demethylation reaction and perhaps a real methyl-arginine demethylase may be found in the JmjC domain family proteins.

The demonstration that JmjC domain proteins function as histone demethylases significantly extends recent findings that LSD1 is an H3-K4 demethylase⁷. First, it reveals a new reaction mechanism that differs from the LSD1 family of proteins in its cofactor requirement

and the reaction products it produces. Second, the LSD1 protein family encompasses only about ten related proteins²³, of which only a few are potential histone demethylases, making it unlikely that these proteins are responsible for modulating the diverse range of histone methylation states²⁴. In contrast, JmjC domain proteins form a large protein family that could potentially satisfy the diverse range of enzymatic requirements to regulate and counteract histone methylation. Third, unlike LSD1, which requires a protonated nitrogen for the demethylation reaction, thus limiting the ability of LSD1 to demethylate trimethyl-lysine, there is no such limitation for JmjC domain-mediated demethylation. Given that JmjC domain-mediated demethylation should have the capacity to demethylate the trimethyl state, it was surprising to us that JHDM1A did not carry out this function. One possible explanation is that the methyl-state specificity of JHDM1A might be regulated by its associated proteins. A precedent exists for this possibility, as the H3-K9 methyltransferase ESET alone can only methylate H3-K9 to the dimethyl state, and requires association with mAM for H3-K9 trimethylation⁶. In a more dramatic example, association of androgen receptor with LSD1 can convert LSD1 from a K4 demethylase to a K9 demethylase²⁵. A second possibility is that the catalytic pocket of JHDM1A may be too small to accommodate K36me3. Consistent with this possibility, a mono-methylase can be converted to a trimethylase by creating a slightly larger catalytic pocket²⁶. A third possibility is that an unshared electron pair on the nitrogen of the ϵ -amino group of lysine is involved in a hydrogen bond within the JHDM1A catalytic site so that the substrate methyl-lysine can be positioned in the catalytic site. In the later two cases, demethylation of trimethyl-lysine may need a JmjC domain with either a larger catalytic pocket or a way of positioning the substrate lysine that does not require the presence of an unshared electron pair on the nitrogen of the ϵ -amino group. A final possibility is that the trimethyl state is permanent and cannot be reversed. Further characterization of JHDM1A, particularly structural studies, should help to explain why it fails to demethylate K36me3.

METHODS

In vitro histone demethylase assay. To prepare for ³H-labelled methyl-histone octamers or oligonucleosome substrates, histone methyltransferase (HMT) assays were performed as previously described¹³. The HMTs used include GST-SET7, GST-G9a, CBP-Set2-Flag, GST-hDOT1L, GST-PRMT1, GST-Suv4-20h1 and the EZH2 complex. After the HMT reaction, the reaction mixtures were dialysed into histone storage buffer (10 mM HEPES-KOH (pH 7.5), 10 mM KCl, 0.2 mM PMSF and 10% glycerol) before being used in histone demethylase assays.

For histone demethylase assays, core histones, nucleosomes (either ³H-labelled or not) or H3-K36 methylated peptide substrates were incubated with protein fractions or purified Flag-JHDM1A in histone demethylation buffer (50 mM HEPES-KOH (pH 8.0), 7–700 μ M Fe(NH₄)₂(SO₄)₂, 1 mM α -ketoglutarate, 2 mM ascorbate) at 37 °C for 1–3 h. Demethylation was analysed by the NASH method, western blotting and mass spectrometry. For detection of ³H-labelled formaldehyde, a modified NASH method²⁷ was used. After TCA precipitation, equal volume of NASH reagent (3.89 M ammonium acetate, 0.1 M acetic acid, 0.2% 2,4-pentanedione) was added into the supernatant and the mixtures were incubated at 37 °C for 50 min before being extracted with equal volume of 1-pentanol. The extracted radioactivity was measured by scintillation counting. For detection of demethylation with peptide substrates, peptides in the reaction mixture were desalted on an RP micro-tip and analysed by MALDI-TOF as described in Supplementary Methods. For detection of histone demethylation using western blotting, demethylation reactions were subjected to western blotting using methyl-specific antibodies.

Methods for purification of HMTs, JHDM1A, Flag-JHDM1A and immunostaining are described in the Supplementary Methods. Constructs and antibodies are also described in the Supplementary Methods.

Received 22 September; accepted 15 November 2005.

Published online 18 December 2005.

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

2. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116**, 2117–2124 (2003).
3. Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).
4. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
5. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
6. Wang, H. *et al.* mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol. Cell* **12**, 475–487 (2003).
7. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
8. Hakimi, M. A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl Acad. Sci. USA* **99**, 7420–7425 (2002).
9. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**, 735–738 (2003).
10. You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc. Natl Acad. Sci. USA* **98**, 1454–1458 (2001).
11. Falnes, P. O., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**, 178–182 (2002).
12. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).
13. Wang, H. *et al.* Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **293**, 853–857 (2001).
14. Carr, S. *et al.* The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell. Proteomics* **3**, 531–533 (2004).
15. Cenciarelli, C. *et al.* Identification of a family of human F-box proteins. *Curr. Biol.* **9**, 1177–1179 (1999).
16. Winston, J. T., Koepp, D. M., Zhu, C., Elledge, S. J. & Harper, J. W. A family of mammalian F-box proteins. *Curr. Biol.* **9**, 1180–1182 (1999).
17. Clissold, P. M. & Ponting, C. P. JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2 β . *Trends Biochem. Sci.* **26**, 7–9 (2001).
18. Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 α . *J. Biol. Chem.* **278**, 1802–1806 (2003).
19. Ayoub, N. *et al.* A novel jmjC domain protein modulates heterochromatization in fission yeast. *Mol. Cell. Biol.* **23**, 4356–4370 (2003).
20. Trewick, S. C., McLaughlin, P. J. & Allshire, R. C. Methylation: lost in hydroxylation? *EMBO Rep.* **6**, 315–320 (2005).
21. Cuthbert, G. L. *et al.* Histone deimination antagonizes arginine methylation. *Cell* **118**, 545–553 (2004).
22. Wang, Y. *et al.* Human PAD4 regulates histone arginine methylation levels via demethyliminination. *Science* **306**, 279–283 (2004).
23. Kubicek, S. & Jenuwein, T. A crack in histone lysine methylation. *Cell* **119**, 903–906 (2004).
24. Wysocka, J., Milne, T. A. & Allis, C. D. Taking LSD1 to a new high. *Cell* **122**, 654–658 (2005).
25. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–439 (2005).
26. Zhang, X. *et al.* Structural basis for the product specificity of histone lysine methyltransferases. *Mol. Cell* **12**, 177–185 (2003).
27. Kleeberg, U. & Klinger, W. Sensitive formaldehyde determination with Nash's reagent and a 'tryptophan reaction'. *J. Pharmacol. Methods* **8**, 19–31 (1982).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Lacomis and G. Boysen for help with mass spectrometry; B. Strahl, T. Jenuwein, Y. Shinkai and A. Verreault for reagents; R. Cao for the EZH2 complex; and R. Klose for critical reading of the manuscript. This work was supported by NIH grants to Y.Z., P.T. and C.H.B. Y.Z. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions Y.Z. designed the experimental strategy and wrote the paper; Y.T. worked out the details of the assay and performed most of the experiments; J.F. analysed the yeast protein; H.E.-B., M.E.W., C.H.B. and P.T. performed mass spectrometric analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.Z. (yi_zhang@med.unc.edu).

Transient radio bursts from rotating neutron stars

M. A. McLaughlin¹, A. G. Lyne¹, D. R. Lorimer¹, M. Kramer¹, A. J. Faulkner¹, R. N. Manchester², J. M. Cordes³, F. Camilo⁴, A. Possenti⁵, I. H. Stairs⁶, G. Hobbs², N. D'Amico^{5,7}, M. Burgay⁵ & J. T. O'Brien¹

The radio sky is relatively unexplored for transient signals¹, although the potential of radio-transient searches is high. This was demonstrated recently by the discovery of a previously unknown type of source^{2,3}, varying on timescales of minutes to hours. Here we report a search for radio sources that vary on much shorter timescales. We found eleven objects characterized by single, dispersed bursts having durations between 2 and 30 ms. The average time intervals between bursts range from 4 min to 3 h with radio emission typically detectable for <1 s per day. From an analysis of the burst arrival times, we have identified periodicities in the range 0.4–7 s for ten of the eleven sources, suggesting origins in rotating neutron stars. Despite the small number of sources detected at present, their ephemeral nature implies a total Galactic population significantly exceeding that of the regularly pulsing radio pulsars. Five of the ten sources have periods >4 s, and the rate of change of the pulse period has been measured for three of them; for one source, we have inferred a high magnetic field strength of 5×10^{13} G. This suggests that the new population is related to other classes of isolated neutron stars observed at X-ray and γ -ray wavelengths⁴.

The eleven sources were detected in a search for isolated bursts of radio emission in data recorded for the Parkes multibeam pulsar survey between January 1998 and February 2002. Figure 1 and Supplementary Fig. 1 show example detections. All bursts from a given source have the same unique value of dispersion measure (DM), or integrated free-electron column density, incontrovertibly distinguishing them from impulsive terrestrial interference.

Since August 2003, all the sources have been reobserved at least nine times at intervals of between one and six months. All have shown multiple bursts, with between four and 229 events detected in total from each object (see Table 1). As far as we can tell from the limited statistics, the density of sources on the sky appears to be greater towards the Galactic plane, with eight of the eleven having $|b| < 2^\circ$ (where b is Galactic latitude). Average rates of detected events range from one every three hours for J1911+00 to one every four minutes for J1819–1458. The bursts (2–30 ms long) have peak 1,400-MHz flux densities which range from 0.1 to 3.6 Jy. These sources are therefore among the brightest radio sources in the Universe after the giant pulses detected from the Crab pulsar and the pulsar B1937+21 (ref. 5).

Periodicity searches that depend on a pulsar's time-averaged emission, including a standard Fourier analysis and a fast-folding algorithm⁶, have been carried out on all survey and follow-up observations, with no periodicities detected using these methods. However, for ten of the sources we have been able to identify a periodicity from the arrival times of the bursts themselves (see Table 2). The 0.4–7 s period range indicates that they are likely to be rotating neutron stars. Most of the periods are quite long; five of

the ten have periods exceeding four seconds, compared with only one in 200 of the known radio pulsar population. As shown in Table 2, for three of the sources we have been able to measure period derivatives using standard pulsar timing techniques⁶ on the individual burst arrival times.

How is the bursting behaviour related to single-pulse behaviour of normal radio pulsars? For most of the sources, the number of detected bursts is not yet sufficient to obtain reliable luminosity distributions. In Supplementary Fig. 2, we show the peak flux density distributions for four objects. For J1317–5759 and J1819–1458, the periodic sources with the greatest number of bursts detected, the non-detection in periodicity searches means that the average peak flux density must be less than 0.5% (for J1819–1458) and 0.8% (for J1317–5759) of the peak flux density of the strongest detected bursts. These objects show power-law tails to their burst amplitude distributions, as seen for giant pulses from the Crab pulsar and pulsar B1937+21 (see, for example, ref. 5). However, all pulsars from which giant pulses have been detected appear to have high values of magnetic field strength at their light cylinder radii⁷. While the Crab pulsar has a magnetic field strength at the light cylinder of 9.3×10^5 G, this value ranges from only 3 to 30 G for these sources, suggesting that the bursts originate from a different emission mechanism.

We therefore conclude that these sources represent a previously unknown population of bursting neutron stars, which we call rotating radio transients (RRATs). In Fig. 2, we show their relationship to other neutron star populations. The long periods of some of the RRATs are similar to the apparently radio-quiet X-ray populations of magnetars⁴ and isolated neutron stars⁸. Additionally, the inferred surface dipole magnetic field of J1819–1458 of 5×10^{13} G is greater than the magnetic field of all but four of the 1,600 known radio pulsars, and is comparable to those of the magnetars^{4,9}. This RRAT is young, with a characteristic age of 117 kyr, smaller than those of 94% of all currently known radio pulsars. The period and magnetic field of J1317–5759 are similar to those of J0720.4–3125, the only radio-quiet, X-ray dim, isolated neutron star with a measured period and period derivative¹⁰.

The RRATs for which we have measured period derivatives show no evidence for binary motion. Likewise, we detect no glitches or other timing abnormalities, although continued monitoring is necessary to gauge the regularity of spin-down rates. In the future, radio polarization data may enable us to constrain the emission mechanism. For the three RRATs with accurate positions, a search of high-energy archives¹¹ reveals no X-ray or γ -ray counterparts.

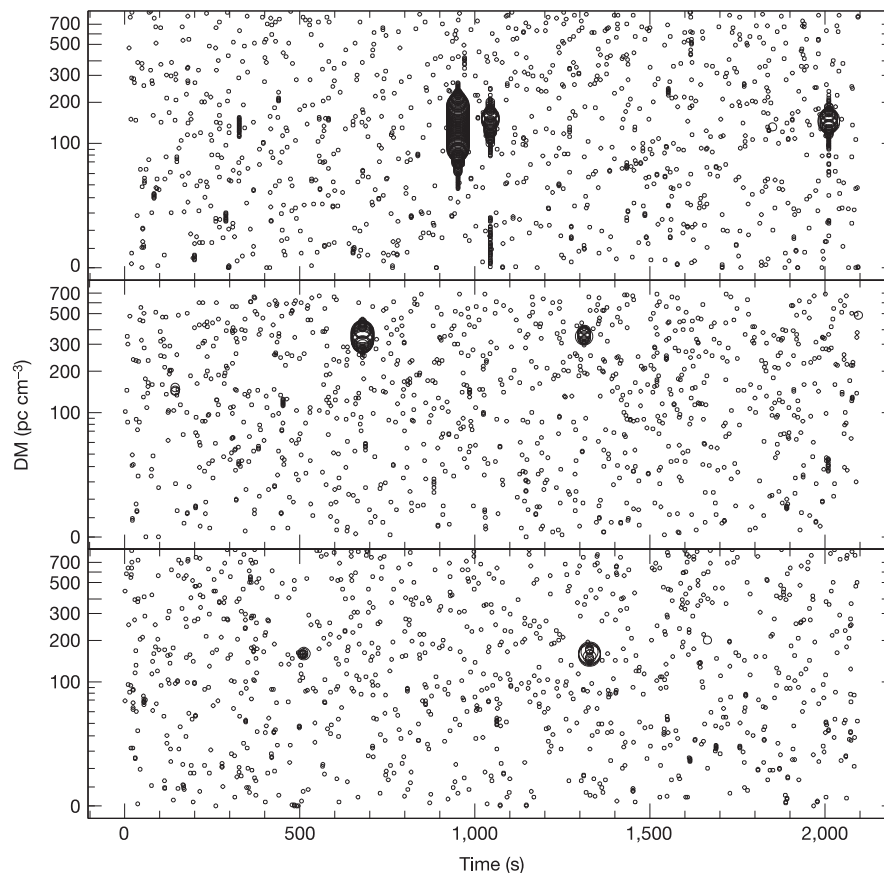
The discovery of this new population results in substantially increased estimates of the total number of Galactic active radio-emitting neutron stars. We detect, on average, one burst for every three hours of observation for J1911+00. The chance of detection

¹Jodrell Bank Observatory, University of Manchester, Macclesfield, Cheshire SK11 9DL, UK. ²Australia Telescope National Facility – CSIRO, PO Box 76, Epping, New South Wales 1710, Australia. ³Astronomy Department and NAIC, Cornell University, Ithaca, New York 14853, USA. ⁴Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, New York 10027, USA. ⁵INAF – Osservatorio Astronomico di Cagliari, Loc. Poggio dei Pini, Strada 54, 09012 Capoterra, Italy. ⁶Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ⁷Dipartimento di Fisica, Università di Cagliari, Strada Provinciale Monserrato-Sestu, km 0.700, 09042 Monserrato, Italy.

Table 1 | Measured and derived parameters for the eleven sources

Name	RA (J2000) (h m s)	Dec (J2000) ($^{\circ}$ ' ")	l ($^{\circ}$)	b ($^{\circ}$)	DM (pc cm $^{-3}$)	D (kpc)	w_{50} (ms)	S_{1400} (mJy)	N_p/T_{obs} (h $^{-1}$)	$N_{\text{det}}/N_{\text{obs}}$
J0848–43	08:48(1)	–43:16(7)	263.4	0.2	293(19)	5.5	30	100	27/19	9/28
J1317–5759	13:17:46.31(7)	–57:59:30.2(6)	306.4	4.7	145.4(3)	3.2	10	1,100	108/24	23/24
J1443–60	14:43(1)	–60:32(7)	316.2	–0.6	369(8)	5.5	20	280	32/41	17/25
J1754–30	17:54(1)	–30:11(7)	359.9	–2.2	98(6)	2.2	16	160	18/30	10/20
J1819–1458	18:19:33.0(5)	–14:58:16(32)	16.0	0.1	196(3)	3.6	3	3,600	229/13	24/24
J1826–14	18:26(1)	–14:27(7)	17.2	–1.0	159(1)	3.3	2	600	18/17	8/12
J1839–01	18:39(1)	–01:36(7)	30.1	2.0	307(10)	6.5	15	100	8/13	1/10
J1846–02	18:46(1)	–02:56(7)	29.7	–0.1	239(10)	5.2	16	250	11/10	5/9
J1848–12	18:48(1)	–12:47(7)	21.1	–5.0	88(2)	2.4	2	450	10/8	5/9
J1911+00	19:11(1)	+00:37(7)	35.7	–4.1	100(3)	3.3	5	250	4/13	4/11
J1913+1333	19:13:17.69(6)	+13:33:20.1(7)	47.5	1.4	175.8(3)	5.7	2	650	66/14	7/10

For each source, we give (left to right) name, right ascension, declination, Galactic longitude, Galactic latitude, DM, inferred distance, average burst duration at 50% of the maximum, peak 1,400-MHz flux density of brightest detected burst, ratio of the total number of bursts detected to the total observation time, and the ratio of the number of observations in which at least one burst was detected to the total number of observations. Estimated 1σ errors are given in parentheses where relevant and refer to the last quoted digit. The mean latitudes and longitudes are comparable to those of the pulsars detected in the Parkes survey. The distances are inferred from their DMs, positions and a model for the Galactic free electron density¹⁶. The mean distance of 4.2 kpc is comparable to that of 5.8 kpc for the pulsars detected in the Parkes survey. The extremely sporadic nature of the bursts makes localization difficult, with most positions known only to within the 1,400-MHz 14-arcmin beam of the Parkes Telescope. For the three sources for which we have measured period derivatives, more accurate positions have been derived through radio timing. Burst durations for each source remain constant, within the uncertainties, and are all much larger than those measured for pulsar giant pulses (see, for example, refs 17, 18).

**Figure 1 | The observational signatures of the new radio transient sources.**

From top to bottom, we show the original detections of J1317–5759, J1443–60 and J1826–14 in the Parkes multibeam survey data. The Parkes survey, which has discovered over 750 radio pulsars¹⁹, used a 13-beam 1,400-MHz cryogenic receiver and covered 1,500 deg² within 5° of the Galactic plane, for longitudes $260^{\circ} < l < 50^{\circ}$ with 250 μ s sampling of a multichannel receiver and 35 min dwell-times on each position²⁰.

Approximately 30% of all pulsars that were detected in the survey using standard periodicity-seeking Fourier techniques were also detected in the burst search. Since radio waves are dispersed by ionized gas in the interstellar medium, the effects of such dispersion have to be removed, and we have therefore used search techniques similar to those described in ref. 21. In short, the 35-min time series were de-dispersed for a number of trial values of dispersion measure (DM). The time series were smoothed by convolution

with boxcars of various widths to increase sensitivity to broadened pulses, with a maximum boxcar width of 32 ms. Because the optimal sensitivity is achieved when the smoothing window width equals the burst width, our sensitivity is lower for burst durations greater than 32 ms. Each of these time series was then searched for any bursts above a threshold of five standard deviations, computed by calculating a running mean and root-mean-square deviation of the noisy time series. All bursts detected above a 5σ threshold are plotted as circles, with size proportional to the signal-to-noise ratio of the detected burst. The abscissa shows arrival time while the ordinate shows the DM. Because of their finite width, intense bursts are detected at multiple DMs and result in vertical broadening of the features. Bursts which are strongest at zero DM and therefore likely to be impulsive terrestrial interference are not shown. In general, these were easily identified by their detection in multiple beams of the 13-beam receiver.

Table 2 | Measured and derived parameters of the 10 sources with measured periods

Name	P (s)	w_{50}/P (%)	Epoch (MJD)	$\dot{P}$ ($10^{-15} \text{ s s}^{-1}$)	B (10^{12} G)	τ_c (Myr)	$\dot{E}$ ($10^{31} \text{ erg s}^{-1}$)
J0848–43	5.97748(2)	0.50	53492	–	–	–	–
J1317–5759	2.6421979742(3)	0.38	53346	12.6(7)	5.83(2)	3.33(2)	2.69(1)
J1443–60	4.758565(5)	0.42	53410	–	–	–	–
J1754–30	0.422617(4)	3.79	53189	–	–	–	–
J1819–1458	4.263159894(6)	0.07	53265	576(1)	50.16(6)	0.1172(3)	24.94(5)
J1826–14	0.7706187(3)	0.26	53587	–	–	–	–
J1839–01	0.93190(1)	1.61	51038	–	–	–	–
J1846–02	4.476739(3)	0.36	53492	–	–	–	–
J1848–12	6.7953(5)	0.03	53158	–	–	–	–
J1913+1333	0.9233885242(1)	0.22	53264	7.87(2)	2.727(4)	1.860(6)	39.4(1)

For each source, we give (left to right) name, barycentric period, average duty cycle (that is, w_{50}/P), the epoch of the period and, if measurable, the period derivative and derived parameters (see below). Periods are derived by calculating the largest common denominator of the differences between the burst arrival times at a given epoch. Given the number of pulses detected per epoch, and the number of epochs for which a periodicity can be measured, we may calculate the probability that the listed period is an integer multiple of the true period. Because of the small number of bursts detected per epoch for J1754–30 and J1848–12, there is a 32% and 16% chance, respectively, that the listed period is actually an integer multiple of the true period. For J1839–01 and J1846–02, this probability is less than 1% and for all others, the probability is less than 0.1%. Assuming that they are rotating neutron stars, the inferred surface dipole magnetic field is calculated as $B = 3.2 \times 10^{19} \sqrt{P\dot{P}}$ G, the characteristic age as $\tau_c = P/2\dot{P}$ and the spin-down luminosity as $\dot{E} = 4\pi^2 I \dot{P} P^{-3}$, where I , the neutron star moment of inertia, is assumed to be 10^{45} g cm^2 (see ref. 6). The duty cycles are generally smaller than those of radio pulsars with similar periods.

within the single 35-min discovery observation was therefore less than 20%, implying that there should be roughly five times the number of similar sources in the same searched volume. Applying a similar analysis to all of the RRATs shows that we expect there to be twice as many sources as we have detected at a similar sensitivity level and sky coverage as for the Parkes survey. This number may be a gross underestimate, however. First, it is very difficult to identify such sources in observations which are contaminated with large amounts of impulsive interference. There may be at least twice as many RRATs that were missed due to this effect. Second, we are only extrapolating to the area covered by the Parkes survey, and the true distribution of these objects is unknown. In addition, because our sensitivity was diminished for burst durations greater than 32 ms, there may be more sources with longer bursts that fell below our detection threshold. Furthermore, previous surveys with observations times

of a few minutes had little chance of detecting such events and most did not include searches for them.

With these caveats in mind, we have carried out a Monte Carlo simulation to provide a first-order estimate of the size of the Galactic RRAT population. The simulation assumes that their spatial distribution follows that derived for the pulsars detected in the Parkes survey¹², that the burst-duration distribution is similar to that observed in the eleven found so far, and, as measured for the pulsar population¹³, that the differential radio luminosity function of an average burst is of the form $d \log N / d \log L = -1$, where N is the number of model sources above a given luminosity $L = Sd^2$, where S is the peak flux density and d is the distance. By calculating the threshold of our survey to model bursts, and generating Monte Carlo realizations, we find that the simulations produce a good match to the observations, but are fairly insensitive to the minimum burst peak luminosity $L_{\min}$ which could plausibly lie in the range 1–100 mJy kpc². To be consistent with the detection of eleven sources reported here, the implied size of the Galactic population of RRATs $N \approx 4 \times 10^5 (L_{\min})^{-1} \times (0.5/f_{\text{on}}) \times (0.5/f_{\text{int}}) \times (0.1/f_{\text{b}})$, where $L_{\min}$ is in units of 10 mJy kpc², f_{on} is the fraction of sources with bursts visible within our 35-min observation, f_{int} is the fraction of bursts not missed owing to interference and f_{b} is the fraction of RRATs whose bursts are beamed towards the Earth. The average beaming fraction for pulsars is roughly 10%, and decreases for longer period pulsars¹⁴. Given the small RRAT duty cycles (see Table 2), our adopted f_{b} is almost certainly a conservative overestimate. An $L_{\min}$ of 10 mJy kpc² is consistent with the lowest peak luminosities observed for the single pulses of known radio pulsars. Assuming that the total Galactic population of active radio pulsars is of the order of 10^5 (see, for example, ref. 15), this discovery increases the current Galactic population estimates by at least several times. We therefore expect the emerging generation of wide-field radio telescopes to discover many more RRATs.

Received 25 August; accepted 15 November 2005.

1. Cordes, J. M., Lazio, T. J. W. & McLaughlin, M. A. The dynamic radio sky. *New Astron. Rev.* **48**, 1459–1472 (2004).
2. Hyman, S. D. *et al.* A powerful bursting radio source towards the Galactic Centre. *Nature* **434**, 50–52 (2005).
3. Kulkarni, S. R. & Phinney, E. S. Astronomy: blasts from the radio heavens. *Nature* **434**, 28–29 (2005).
4. Woods, P. M. & Thompson, C. in *Compact Stellar X-Ray Sources* (eds Lewin, W. & van der Klis, M.) (Cambridge Univ. Press, in the press).
5. Cordes, J. M., Bhat, N. D. R., Hankins, T. H., McLaughlin, M. A. & Kern, J. The brightest pulses in the Universe: multifrequency observations of the Crab pulsar's giant pulses. *Astrophys. J.* **612**, 375–388 (2004).
6. Lorimer, D. R. & Kramer, M. *Handbook of Pulsar Astronomy* (Cambridge Univ. Press, 2005).
7. Johnston, S. & Romani, R. W. Giant pulses from PSR B0540–69 in the Large Magellanic Cloud. *Astrophys. J.* **590**, L95–L98 (2003).

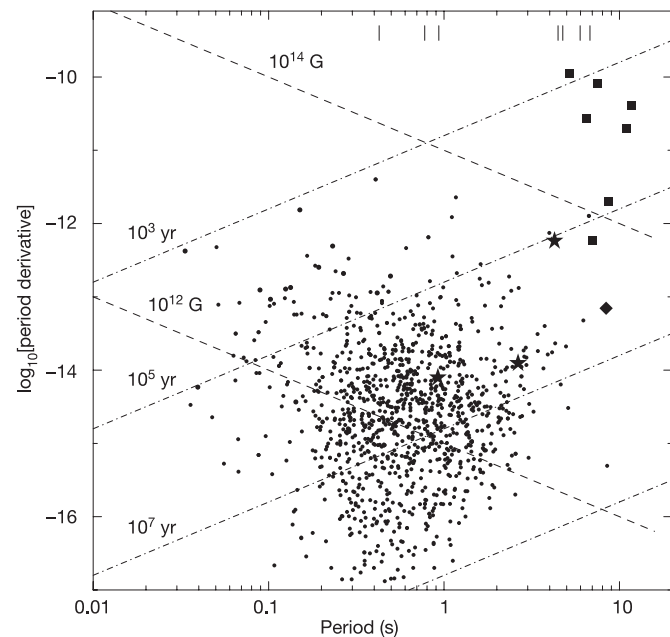


Figure 2 | The rotational properties of neutron stars summarized in a P – $\dot{P}$ diagram. The rotational period derivative ($\dot{P}$) is plotted against period (P) for pulsars (dots), magnetars (squares), the one radio-quiet, X-ray dim, isolated neutron star with a measured period and period derivative (diamond), and the three RRATs having measured periods and period derivatives (stars). The vertical lines at the top of the plot mark the periods of the other seven sources in Table 2. Dashed lines indicate the loci of constant values of characteristic age and inferred surface dipole magnetic field strength.

8. Haberl, F. AXPs and X-ray-dim isolated neutron stars: recent XMM-Newton and Chandra results. *Adv. Space Res.* **33**, 638–644 (2004).
9. McLaughlin, M. A. *et al.* PSR J1847-0130: a radio pulsar with magnetar spin characteristics. *Astrophys. J.* **591**, L135–L138 (2003).
10. Kaplan, D. L. & van Kerkwijk, M. H. A coherent timing solution for the nearby isolated neutron star RX J0720.4-3125. *Astrophys. J.* **628**, L45–L48 (2005).
11. NASA's HEASARC: Archive. (<http://heasarc.gsfc.nasa.gov/docs/archive.html>) (2005).
12. Lorimer, D. R. in *Young Neutron Stars and Their Environments* (eds Camilo, F. & Gaensler, B. M.) **105**, 105–112 (IAU Symp. 218, Astronomical Society of the Pacific, San Francisco, 2004).
13. Lyne, A. G. *et al.* The Parkes Southern Pulsar Survey II. Final results and population analysis. *Mon. Not. R. Astron. Soc.* **295**, 743–755 (1998).
14. Tauris, T. M. & Manchester, R. N. On the evolution of pulsar beams. *Mon. Not. R. Astron. Soc.* **298**, 625–636 (1998).
15. Vranesevic, N. *et al.* Pulsar birthrates from the Parkes Multibeam Survey. *Astrophys. J.* **617**, L139–L142 (2004).
16. Cordes, J. M. & Lazio, T. J. W. NE2001. I. A new model for the Galactic distribution of free electrons and its fluctuations. Preprint at (<http://arxiv.org/astro-ph/0207156>) (2002).
17. Soglasnov, V. A. Giant pulses from PSR B1937+21 with widths less than 15 nanoseconds and T greater than 5×10^{39} K, the highest brightness temperature observed in the Universe. *Astrophys. J.* **616**, 439–451 (2004).
18. Hankins, T. H., Kern, J. S., Weatherall, J. C. & Eilek, J. A. Nanosecond radio bursts from strong plasma turbulence in the Crab pulsar. *Nature* **422**, 141–143 (2003).
19. Hobbs, G. *et al.* The Parkes multibeam pulsar survey—IV. Discovery of 180 pulsars and parameters for 281 previously known pulsars. *Mon. Not. R. Astron. Soc.* **352**, 1439–1472 (2004).
20. Manchester, R. N. *et al.* The Parkes multi-beam pulsar survey—I. Observing and data analysis systems, discovery and timing of 100 pulsars. *Mon. Not. R. Astron. Soc.* **328**, 17–35 (2001).
21. Cordes, J. M. & McLaughlin, M. A. Searches for fast radio transients. *Astrophys. J.* **596**, 1142–1154 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Parkes radio telescope is part of the Australia Telescope, which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. A.P. and N.D'A. acknowledge financial support from the Italian Ministry of University and Research (MIUR). F.C. is supported by NSF, NASA and NRAO. D.R.L. is a University Research Fellow funded by the Royal Society. I.H.S. holds an NSERC UFA and is supported by a discovery grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.McL. (Maura.McLaughlin@manchester.ac.uk).

Iron meteorites as remnants of planetesimals formed in the terrestrial planet region

William F. Bottke¹, David Nesvorný¹, Robert E. Grimm¹, Alessandro Morbidelli² & David P. O'Brien²

Iron meteorites are core fragments from differentiated and subsequently disrupted planetesimals¹. The parent bodies are usually assumed to have formed in the main asteroid belt, which is the source of most meteorites. Observational evidence, however, does not indicate that differentiated bodies or their fragments were ever common there. This view is also difficult to reconcile with the fact that the parent bodies of iron meteorites were as small as 20 km in diameter^{2,3} and that they formed 1–2 Myr earlier than the parent bodies of the ordinary chondrites^{4–6}. Here we show that the iron-meteorite parent bodies most probably formed in the terrestrial planet region. Fast accretion times there allowed small planetesimals to melt early in Solar System history by the decay of short-lived radionuclides (such as ²⁶Al, ⁶⁰Fe)^{7–9}. The protoplanets emerging from this population not only induced collisional evolution among the remaining planetesimals but also scattered some of the survivors into the main belt, where they stayed for billions of years before escaping via a combination of collisions, Yarkovsky thermal forces, and resonances¹⁰. We predict that some asteroids are main-belt interlopers (such as (4) Vesta). A select few may even be remnants of the long-lost precursor material that formed the Earth.

In this view, inner Solar System planetesimals or their fragments, presumed to be the parent bodies of most iron meteorites, are scattered into the main asteroid belt early in its history. To investigate this, we used numerical simulations to track thousands of massless test bodies (planetesimals) evolving amid a swarm of Moon- to Mars-sized planetary embryos spread between 0.5–3.0 AU (see Fig. 1 for computational details). Figure 1 shows a representative snapshot of a thousand of our test bodies after 10 Myr of evolution, with subgroups having initial semimajor axis a values between 0.5–1.0 AU, 1.0–1.5 AU and 1.5–2.0 AU. We find that planetary embryo perturbations increase the mean displacement of particles from the centre of each group over time; for test bodies that maintain eccentricity $e < 0.3$ and inclination $i < 15^\circ$, we find, for each subgroup, $\langle \Delta a \rangle \approx 0.2$ AU at 4 Myr and 0.3–0.4 AU at 10 Myr. These results are consistent with the observed semimajor axis spread of large S- and C-type asteroids in the main belt¹³.

The most intriguing part of Fig. 1, however, are the outliers who enter the main-belt zone through a combination of resonant interactions and close encounters with planetary embryos. Figure 2 shows that nearly 0.01–0.1%, 1% and 10% of the particles respectively started with $a = 0.5$ –1.0 AU, 1.0–1.5 AU, and 1.5–2.0 AU achieve main-belt orbits. Once there, these objects are dynamically indistinguishable from the rest of the main-belt population; while many may be ejected over time via interactions with planet embryos, resonances, and so on^{14,15}, the proportion of interloper to indigenous material in the main belt should stay the same. Figure 1 also shows that much of this material is delivered to the inner main belt, where meteoroids are dynamically most likely to reach Earth (ref. 16; see also Supplementary

Discussion). We infer from these results that interloper material should be an important component in the meteorite collection.

If planetesimal material from the terrestrial planet region were actually in the main belt, we can guess at its nature by examining the main-belt population. Observations show a broad-scale taxonomic stratification among large main-belt asteroids, with S-type asteroids, believed to be analogous to metamorphosed but unmelted ordinary chondrites, dominating the inner main belt and C-type asteroids, believed to be analogous to more primitive carbonaceous chondrites, dominating the outer main belt^{13,17}. This trend, if followed inward towards the Sun, implies that inner Solar System planetesimals experienced significantly more heating than S- and C-type asteroids, with the most plausible planetesimal heat source being radionuclides like ²⁶Al and ⁶⁰Fe (refs 7, 8, 9). Because the half-lives of these isotopes are only 0.73 Myr and 1.5 Myr, respectively, bodies that accrete quickly stand the best chance of undergoing differentiation. Although precise accretion timescales across the inner Solar System are unknown, modelling work suggests they vary with swarm density and a , such that accretion timescales increase with increasing heliocentric distance (at least until the so-called ‘snowline’ is reached)^{18–20}. Accordingly, if main-belt interlopers are derived from regions closer to the Sun, their shorter accretion times would lead to more internal heating¹⁷ and thus they would probably look like heavily metamorphosed or differentiated asteroids.

At this point, a plausible connection can be made between our putative interlopers and iron meteorites. Cooling rate and textural data from irons indicate that most come from the cores of small differentiated asteroids (diameter $D \approx 20$ –200 km; refs 2, 3); very few are thought to be impact melts or fragments from larger differentiated bodies (for example, the $D = 530$ km asteroid (4) Vesta)^{2,21}. Isotopic chronometers also indicate that core formation among iron meteorite parent bodies occurred 1–2 Myr before the formation of the ordinary chondrite parent bodies^{4,5,6}. The paradox is that if small asteroids differentiated in the main belt at such early times, it would be reasonable to expect larger bodies forming near the same locations to have differentiated as well (ref. 17; see also Supplementary Discussion). Hence, if iron meteorites are indigenous to the main belt, large numbers of differentiated bodies and their fragments should reside there today. This is not observed. Instead, we argue that a more natural formation location for most iron-meteorite parent bodies is the terrestrial planet region, where accretion occurs quickly and thus differentiation is more likely to occur among small bodies¹⁷. A small fraction of this material would then be scattered into the main belt by interactions with planetary embryos.

The paucity of intact differentiated asteroids (or their fragments) in the main belt today, particularly in the inner main belt where numerous meteorites come from¹⁶, is an important constraint for our scenario. For example, despite extensive searches²², only one asteroid is known to sample the crust of a non-Vesta but Vesta-like

¹Southwest Research Institute, 1050 Walnut St, Suite 400, Boulder, Colorado 80302, USA. ²Observatoire de la Côte d’Azur B.P. 4229, 06034 Nice Cedex 4, France.

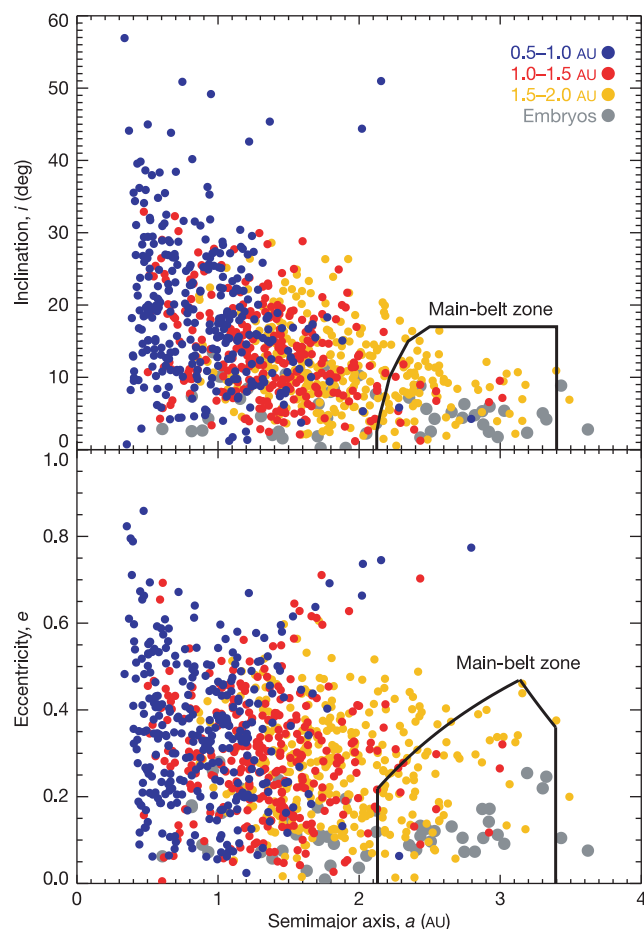


Figure 1 | A snapshot of inner Solar System planetesimals and planetary embryos after 10 Myr of dynamical evolution. The starting conditions and methods used were the same as in ref. 11. We assumed that the jovian planets, if they existed, had a negligible effect on the early dynamical history of these bodies. Gas drag and dynamical friction between the embryos and planetesimals were neglected. Tests indicate that these approximations, while imperfect, mainly affect the details of our model rather than the overall story (see Supplementary Discussion). Four sets of 100 embryos (grey dots) were distributed over 0.5–3.0 AU such that their surface density varied as the heliocentric distance $r^{-3/2}$, with 8.0 g cm^{-2} at 1 AU. Each embryo was 0.04 Earth masses. Their initial (e, i) values were chosen randomly from a uniform distribution $e = \{0.5, 5.0\} (a/r_H)$ and $i = 0.5 (e)$, where r_H is an embryo's Hill radius. The planetesimals were given uniform a between 0.5–2.0 AU and random e, i according to a Rayleigh distribution ($i = 0.5 (e)$, with $\langle e \rangle = 0.02$). The blue, red and yellow dots show what happens to 1,000 planetesimals started with 0.5–1.0 AU, 1.0–1.5 AU, and 1.5–2.0 AU, respectively. The black line is the location of the main asteroid belt ($2.0 \text{ AU} < a < 3.5 \text{ AU}$, e is such that the objects do not cross the orbits of Mars or Jupiter, i is below the ν_6 resonance for $2.0 \text{ AU} < a < 2.5 \text{ AU}$, and $i < 17^\circ$ for $2.5 \text{ AU} < a < 3.5 \text{ AU}$)¹². Numerous planetesimals (one blue and several red/yellow) were driven into the main belt by gravitational interactions with embryos, with the highest concentration in the inner main-belt region.

differentiated asteroid: (1459) Magnya, a $D = 20\text{--}30 \text{ km}$ V-type asteroid located in the outer main belt²³ (though see also ref. 24 and Supplementary Discussion). Moreover, main-belt asteroid families, which contain fragments produced by the disruption of over fifty $D \approx 10\text{--}400 \text{ km}$ asteroids, show little spectroscopic evidence that their parent bodies were heated enough to produce a distinct core, mantle and crust (other than those associated with Vesta)²⁵. These data, which suggest that differentiated material from small parent bodies is rare in the main belt, must be reconciled with the following facts: (1) iron meteorites represent over two-thirds of

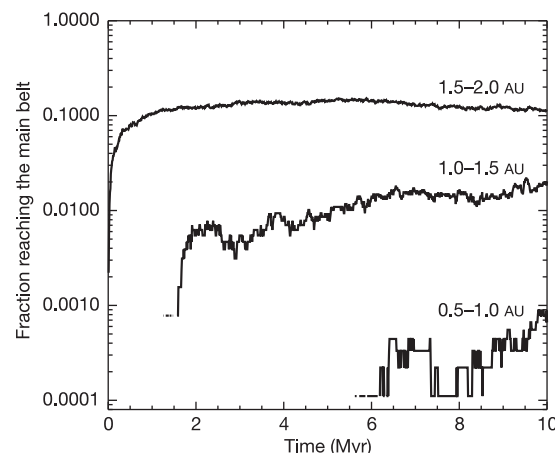


Figure 2 | The fraction of inner Solar System planetesimals scattered into the main-belt zone by gravitational interactions with planetary embryos. Computational details are given in Fig. 1. The curves were generated by tracking 17,000 test bodies for 10 Myr across four planetary embryo simulations. To compute accurate statistics, $>60\%$ of the test bodies were started in the 0.5–1.0 AU zone. The remainder were equally distributed in the 1.0–1.5 AU and 1.5–2.0 AU zones. The proportion of test bodies reaching the main belt from the 1.5–2.0 AU zone is $\sim 10\%$ after 1 Myr of evolution. This value quickly reaches a steady state and remains that way for the remainder of the runs. For the 1.0–1.5 AU zone, 0.8–2% are injected into the main belt after a longer delay of 2 Myr, while for 0.5–1.0 AU we find $>0.01\text{--}0.1\%$ after 6 Myr. Thus, it is plausible that the current main belt contains samples from the feeding zones of Mercury, Venus, Earth and Mars, with the limiting factor being the formation times of the terrestrial and jovian planets.

the unique parent bodies sampled among all meteorites¹ and (2) iron meteorites were probably extracted from the cores of $D \approx 20\text{--}200 \text{ km}$ parent bodies through catastrophic impacts^{2,3}.

We investigated this apparent contradiction by modelling the impact history of inner Solar System planetesimals using a well-tested collision evolution code^{26,27} (see Fig. 3 for computational details). Figure 3 shows the fraction of $D = 20, 100$ and 500 km planetesimals that survive intact between $0.5 \text{ AU} < a < 2.0 \text{ AU}$ as a function of time. Ideally, these results should be coupled to thermal models describing the minimal size needed for differentiation in each zone as a function of time. This cannot be done, however, until accretion times are better quantified. For this reason, our thermal modelling results are only used to guide the discussion below.

In the 0.5–1.5 AU zone, most $D = 20\text{--}100 \text{ km}$ planetesimals disrupt after a few Myr. Because the break-up of a single planetesimal can produce millions of fragments, however, some of this material should be scattered into the main belt (Fig. 2). Accordingly, we predict that many iron meteorites come from these planetesimals: they form close to the Sun, so they are likely to be differentiated, and very few survive intact, explaining the paucity of small but intact differentiated bodies in the main belt. We note that larger $D = 500 \text{ km}$ planetesimals from this zone, although more difficult to disrupt, are limited in number, such that none are likely to survive the dynamical events that depleted the main belt of much of its population early in its history^{14,15,26,27}. The surviving remnants of this differentiated population are therefore more likely to be fragments than intact objects.

Alternatively, 1.5–2.0 AU planetesimals are increasingly likely to both survive comminution and be scattered into the main belt (Figs 2 and 3). Their longer accretion times, however, mean that only the largest, most insulated ones differentiate. Thus, $D < 100 \text{ km}$ interlopers from this zone are more likely to resemble heavily metamorphized S-type and E-type asteroids than differentiated bodies. Interestingly, these heating trends may help us deduce where Vesta

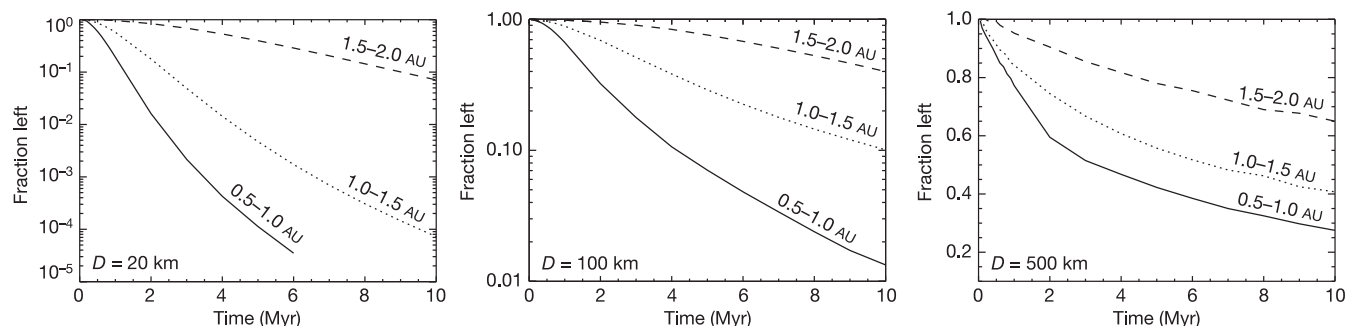


Figure 3 | The fraction of inner Solar System planetesimals that survive the first 10 Myr of collisional evolution. Using an established code^{26,27}, we computed how various input size frequency distributions (SFDs) started at 0.5–1.0 AU, 1.0–1.5 AU, and 1.5–2.0 AU undergo comminution as a function of time. The collision probabilities (P_i) and impact velocities (V_{imp}) of the planetesimals striking one another were computed using data from Figs 1 and 2 (ref. 28). Typical values for our three zones were $P_i \approx 75$, 45 and $17 \times 10^{-18} \text{ km}^{-2} \text{ yr}^{-1}$ and $V_{\text{imp}} \approx 12$, 10 and 8 km s^{-1} , respectively. The input SFD for each zone was assumed to follow the same trends as those determined for the primordial main belt^{14,26,27}. Large planetesimals (diameter $100 < D < 1,000 \text{ km}$) were given a differential power-law index $q \approx -4.5$, the same as that observed in the main belt (and Kuiper belt). The

number of $D > 100 \text{ km}$ bodies was set to ~ 200 times the current main-belt population. Smaller planetesimals ($D < 100 \text{ km}$) were given a shallow initial slope ($q = -1.2$). The results shown here focus on $D = 20$, 100 and 500 km objects. We found that $D = 20 \text{ km}$ planetesimals disrupt quickly enough between 0.5–1.5 AU that only their fragments are likely to reach the main-belt zone. Intact $D = 100 \text{ km}$ planetesimals from 0.5–1.5 AU have a better chance of reaching the main belt, but few then go on to survive the dynamical events that deplete the primordial main belt of its material^{14,15}. Relatively few $D = 500 \text{ km}$ bodies disrupt in any zone. Their limited numbers, however, imply that only those formed in the 1.5–2.0 AU zone are likely to be in the main belt today.

formed. If $D = 500 \text{ km}$ bodies were close to the minimum size needed to differentiate in the main-belt region ($a > 2.0 \text{ AU}$), numerous smaller bodies ($D \leq 500 \text{ km}$) should have differentiated in the adjacent 1.5–2.0 AU zone; according to Fig. 2, many should have been injected into the main belt. We consider it unlikely that collisional and dynamical processes would have eliminated all the evidence. Alternatively, if $D = 500 \text{ km}$ bodies were close to the minimal size needed for differentiation in the 1.5–2.0 AU zone, the paucity of differentiated material in the main belt is more naturally explained, with Vesta perhaps the lone differentiated survivor from that zone.

If samples of crust, mantle and core material from differentiated planetesimals were implanted in the main belt early in its history, why do we find so few olivine and basaltic meteorites from sources other than Vesta? To examine this issue, we tracked the evolution of a hypothetical population of mantle-type material in the inner main belt over the last 4 Gyr in response to comminution and dynamical (Yarkovsky) depletion (see ref. 29 and Supplementary Discussion for details). Our results indicate that there are insufficient A-type asteroids in the inner main belt to keep a large flux of olivine meteoroids continually replenished over several Gyr through a collisional cascade. Thus, while olivine-rich A-type asteroids clearly exist in the inner main belt, they are statistically unlikely to produce a significant number of present-day meteorites. Iron meteoroids, on the other hand, have several advantages over stones: (1) their cosmic-ray exposure ages suggest they are roughly an order of magnitude stronger than stones³⁰, meaning they are less susceptible to comminution and are more likely to survive atmospheric entry, and (2) their high thermal conductivities mean they evolve more slowly by the Yarkovsky effect than do stones¹⁰. Together, these factors mean that the population of small iron asteroids in the inner main belt has probably experienced minimal changes over the last $\sim 4 \text{ Gyr}$.

These results have important implications for asteroid and meteorite studies. For example, it is plausible that some iron meteorites are remnants of the precursor material comprising the terrestrial planets. Hence, by locating and studying crust or mantle fragments associated with these objects in the main belt, we may be able to deduce the composition of the primordial Earth. Note that observable remnants of Earth's precursor material may still be located in the inner main belt, although extensive spectroscopic surveys will be needed to identify them among the background population (see Supplementary Discussion). Our model may also help to explain some curious inconsistencies in the main belt. For example, the

largest main-belt asteroid Ceres ($D = 930 \text{ km}$) and Vesta have very different compositions and thermal histories, despite only being separated by $\sim 0.4 \text{ AU}$. As described above, one possible explanation for the difference is that Vesta is a main-belt interloper. A second and equally likely possibility, however, is that Ceres formed far from the Sun and was scattered inward by planetary embryos. We conclude that the main belt may be the last, best place to look for the long-lost precursors of many Solar System planets.

Received 14 September; accepted 12 December 2005.

- Burbine, T. H., McCoy, T. J., Meibom, A., Gladman, B. & Keil, K. in *Asteroids III* (eds Bottke, W. F. et al.) 653–667 (Univ. Arizona Press, Tucson, 2002).
- Mittlefehldt, D. W., McCoy, T. J., Goodrich, C. A. & Kracher, A. in *Planetary Materials* (ed. Papike, J. J.), *Rev. Mineral.* **36**, 4–195 (1998).
- Chabot, N. L. & Haack, H. in *Meteorites and the Early Solar System II* (eds Lauretta, D. S. & McSween, H. Y.) (Univ. Arizona Press, Tucson, in the press).
- Kleine, T., Mezger, K., Palme, H. & Scherer, E. Tungsten isotopes provide evidence that core formation in some asteroids predates the accretion of chondrite parent bodies. *Lunar Planet. Sci. Conf.* **36**, 1431–1432 (2005).
- Baker, J., Bizzarro, M., Wittig, N., Connolly, J. & Haack, H. Early planetesimal melting from an age of 4.5662 Gyr for differentiated meteorites. *Nature* **436**, 1127–1131 (2005).
- Bizzarro, M., Baker, J. A., Haack, H. & Lundgaard, K. L. Rapid timescales for accretion and melting of differentiated planetesimals inferred from ^{26}Al – ^{26}Mg chronometry. *Astrophys. J.* **632**, L41–L44 (2005).
- Russell, S. S. et al. Evidence for widespread ^{26}Al in the solar nebula and constraints for nebula time scales. *Science* **273**, 757–762 (1996).
- Srinivasan, G., Goswami, J. N. & Bhandari, N. ^{26}Al in eucrite Piplia Kalan: Plausible heat source and formation chronology. *Science* **284**, 1348–1350 (1999).
- Sanders, I. S. & Taylor, G. J. in *Chondrites and the Protoplanetary Disk* (eds Krot, A. N. et al.) 915–932 (ASP Conf. Series 341, Astron. Soc. Pacific, San Francisco, 2005).
- Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Brož, M. in *Asteroids III* (eds Bottke, W. F. et al.) 395–408 (Univ. Arizona Press, Tucson, 2002).
- Levison, H. F. & Agnor, C. The role of giant planets in terrestrial planet formation. *Astron. J.* **125**, 2692–2713 (2003).
- Bottke, W. F. et al. Debaised orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* **156**, 399–433 (2002).
- Gradie, J. & Tedesco, E. Compositional structure of the asteroid belt. *Science* **216**, 1405–1407 (1982).
- Petit, J., Morbidelli, A. & Chambers, J. The primordial excitation and clearing of the asteroid belt. *Icarus* **153**, 338–347 (2001).
- Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469 (2005).
- Gladman, B. J. et al. Dynamical lifetimes of objects injected into asteroid belt resonances. *Science* **277**, 197–201 (1997).
- Grimm, R. E. & McSween, H. Y. Heliocentric zoning of the asteroid belt by aluminum-26 heating. *Science* **259**, 653–655 (1993).

18. Weidenschilling, S. J. The distribution of mass in the planetary system and solar nebula. *Astrophys. Space Sci.* **51**, 153–158 (1977).
19. Greenberg, R., Hartmann, W. K., Chapman, C. R. & Wacker, J. F. Planetesimals to planets—Numerical simulation of collisional evolution. *Icarus* **35**, 1–26 (1978).
20. Stevenson, D. J. & Lunine, J. I. Rapid formation of Jupiter by diffuse redistribution of water vapor in the solar nebula. *Icarus* **75**, 146–155 (1988).
21. Scott, E. R. D. in *Asteroids III* (eds Bottke, W. F. et al.) 697–709 (Univ. Arizona Press, Tucson, 2002).
22. Bus, S. J. & Binzel, R. P. Phase II of the small main-belt asteroid spectroscopic survey: The observations. *Icarus* **158**, 106–145 (2002).
23. Lazzaro, D. et al. Discovery of a basaltic asteroid in the outer main belt. *Science* **288**, 2033–2035 (2000).
24. Sunshine, J. M. et al. High-calcium pyroxene as an indicator of igneous differentiation in asteroids and meteorites. *Meteorit. Planet. Sci.* **39**, 1343–1357 (2004).
25. Cellino, A., Bus, S. J., Doressoundiram, A. & Lazzaro, D. in *Asteroids III* (eds Bottke, W. F. et al.) 632–643 (Univ. Arizona Press, Tucson, 2003).
26. Bottke, W. F. et al. The fossilized size distribution of the main asteroid belt. *Icarus* **175**, 111–140 (2005).
27. Bottke, W. F. et al. Linking the collisional evolution of the main belt to its dynamical excitation and depletion. *Icarus* **179**, 63–94 (2005).
28. Bottke, W. F., Nolan, M. C., Greenberg, R. & Kolvoord, R. A. Velocity distributions among colliding asteroids. *Icarus* **107**, 255–268 (1994).
29. Bottke, W. F. et al. in *Dynamics of Population of Planetary Systems (IAU Colloquium 197, Belegarde, 357–376)* (eds Knezevic, Z. & Milani, A.) (Cambridge Univ. Press, Cambridge, 2005).
30. Eugster, O. Cosmic-ray exposure ages of meteorites and lunar rocks and their significance. *Chemie Erde* **63**, 3–30 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Bus, J. Chambers, D. Durda, M. Gounelle, H. Haack, H. Levison, T. McCoy, D. Mittlefehldt, E. Scott, J. Sunshine, D. Vokrouhlicky and M. Zolensky for discussions and comments. The project was supported by NASA's Origins of Solar System and Planetary Geology and Geophysics programmes.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.F.B. (bottke@boulder.swri.edu).

A spin triplet supercurrent through the half-metallic ferromagnet CrO₂

R. S. Keizer¹, S. T. B. Goennenwein^{1†}, T. M. Klapwijk¹, G. Miao^{2,3}, G. Xiao³ & A. Gupta²

In general, conventional superconductivity should not occur in a ferromagnet, though it has been seen in iron under pressure¹. Moreover, theory predicts that the current is always carried by pairs of electrons in a spin singlet state², so conventional superconductivity decays very rapidly when in contact with a ferromagnet, which normally prohibits the existence of singlet pairs. It has been predicted that this rapid spatial decay would not occur if spin triplet superconductivity could be induced in the ferromagnet^{3,4}. Here we report a Josephson supercurrent through the strong ferromagnet CrO₂, from which we infer that it is a spin triplet supercurrent. Our experimental set-up is different from those envisaged in the earlier predictions, but we conclude that the underlying physical explanation for our result is a conversion from spin singlet pairs to spin triplets at the interface. The supercurrent can be switched with the direction of the magnetization, analogous to spin valve transistors, and therefore could enable magnetization-controlled Josephson junctions.

In our experiment we realized a sample (Fig. 1b) in which two *s*-wave superconductors, made out of NbTiN (ref. 5), are coupled by the conducting ferromagnet CrO₂, a material well known from magnetic recording tapes⁶. Using electron-beam lithography, sputtering and lift-off, two 'T'-shaped NbTiN electrodes with a relatively large mutual distance of 0.3–1 μm are patterned on top of the CrO₂ (Fig. 1c). On cooling the sample to temperatures between 1 and 10 K, we find that the current between the two electrodes, which can only pass through the ferromagnetic CrO₂ film, is a supercurrent (Fig. 2a). We also find that with increasing temperatures the maximum supercurrent, I_c , decreases, and disappears at a temperature comparable to the superconducting transition temperature, T_c , of the thin NbTiN film (Fig. 2b). The observation of a supercurrent through a ferromagnet has been reported before^{7,8}, but only for very weak ferromagnets and over significantly shorter distances.

The ferromagnet we use, CrO₂, is a so-called half-metallic ferromagnet⁶. The electronic transport is metallic for the spin-up electrons, while it is insulating for the spin-down electrons, a property which is entirely due to the band structure of the material, schematically shown in Fig. 1a. Some details of the electronic structure are still under debate, but CrO₂ is assumed to be a self doped double exchange ferromagnet with a gap (of $\sim 2\text{ eV}$) in the spin-down density-of-states at the Fermi level⁹. The material has a Curie temperature $T_{\text{Curie}} \approx 390\text{ K}$ and is metallic at low temperatures¹⁰, with a resistivity of about $8.9\ \mu\Omega\text{ cm}$ at 1.6 K. It has been experimentally demonstrated that, as expected, the spin polarization is close to 100% (refs 11, 12). The saturation magnetization is equal to two Bohr magnetons ($2\mu_B$) per unit cell. Important for our experiments is that the magnetic behaviour (switching/rotation of the magnetization direction) has been shown to be single-domain-like, even for macroscopic films¹³.

From magnetotransport measurements at temperature $T = 4.2\text{ K}$ (in which the resistance of the film is measured as a function of an external, in-plane magnetic field), we find the CrO₂ films to have a biaxial (cubic) magneto-crystalline symmetry in the plane of the film. This biaxial character appears as two switches in the resistance as a function of field, which enabled us to identify the directions of the easy axes of the ferromagnet: 30° , 150° , 210° and 330° to the crystallographic *c* axis. As shown in Fig. 1d, the junction is aligned along this axis, while the *a* axis and *b* axis are out-of-plane and in-plane, respectively.

In conventional superconductors—such as NbTiN used here—electrons are paired in so-called singlet Cooper pairs. In singlet pairs, an electron with spin up is paired with another electron with spin

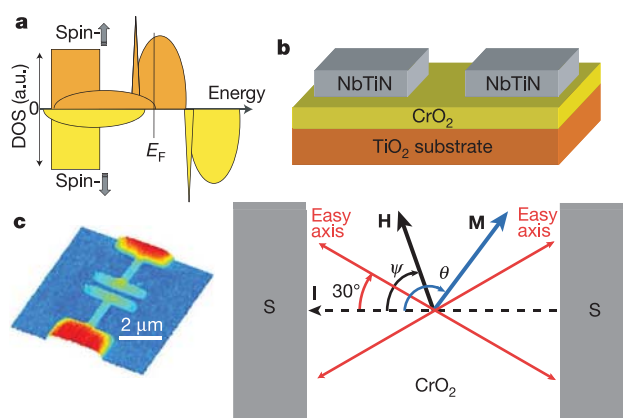


Figure 1 | Basic aspects of the experimental system. **a**, Simplified view of the spin dependent density-of-states (DOS) of CrO₂. At the Fermi level, there is a gap in the DOS for spin-down, while the spin-up band is metallic, leading to a fully spin-polarized conductor. The absence of spin-down states rules out spin-flip scattering in the transport. E_F Fermi energy. **b**, Schematic illustration of the studied devices. The half-metallic CrO₂ (100) single crystal thin film (100 nm thick) is epitaxially grown on top of an (insulating) TiO₂ (100) substrate by chemical vapour deposition²³. Then, patterns are defined in an organic resist mask with conventional electron beam lithography. Before the deposition of the *s*-wave superconductor NbTiN (ref. 5) contacts, the surface not covered by the mask is sputter-cleaned with an Ar plasma to remove the natural oxide Cr₂O₃ terminating the surface in order to obtain high quality contacts to the CrO₂. Note that this fabrication procedure precludes spurious connections between the NbTiN electrodes, as independently confirmed by scanning electron microscopy. **c**, Scanning electron micrograph of a typical final device. **d**, Illustration of the alignment of the current direction with respect to the magnetization axes. ψ is the angle of the applied magnetic field *H* with respect to the current direction *I*, and θ is the direction of the magnetization *M*.

¹Kavli Institute of NanoScience, Faculty of Applied Sciences, Delft University of Technology, 2628 CJ, Delft, The Netherlands. ²MINT Center, University of Alabama, Tuscaloosa, Alabama 35487, USA. ³Physics Department, Brown University, Providence, Rhode Island 02912, USA. [†]Present address: Walther-Meißner-Institut, Bayerische Akademie der Wissenschaften, Walther-Meißner-Straße 8, D-85748, Garching, Germany.

down. When two superconductors are coupled through a normal metal, a supercurrent will flow when the thickness of the normal layer is less than, or of the order of, the normal metal coherence length $\xi_N = \sqrt{\hbar D / k_B T}$, with Planck's constant $\hbar$, the diffusion constant for elastic scattering D , and Boltzmann's constant k_B . In normal metals, $\xi_N \approx 100$ nm is a measure for the length over which a Cooper pair loses its coherence, and is insensitive to the spin of the electrons forming the Cooper pair. If the normal metal is replaced by a conducting ferromagnet, as in our experiment, the electrons sense the magnetization and are pulled apart in energy depending on their spin orientation. ξ_N must then be replaced by $\xi_F \approx \sqrt{\hbar D / k_B T_{\text{Curie}}}$. As typically $T \ll T_{\text{Curie}}$, ξ_F in a ferromagnet is short, of the order of 1 nm. The reason is that a singlet Cooper pair (just as a normal metal) is symmetric for the spin directions. In contrast, a ferromagnet likes to have all spins pointing in the same direction. Several recent experiments have confirmed this short-length-scale picture^{7,8}.

However, Bergeret, Volkov and Efetov³ have demonstrated theoretically that coherent triplet Cooper pairs—that is, pairs of electrons with both spins in parallel—can be induced in a ferromagnet in close proximity to a conventional singlet pair superconductor, given suitable conditions¹⁴. They also predict that the coherence length will be equal to ξ_N of a normal metal, hence the name 'long range proximity effect'. These authors furthermore show⁴ that triplet correlations are, unlike the anomalous order parameters in ³He (ref. 15) and unconventional superconductors^{16,17}, robust against impurity scattering. There are some indications for the existence of these long-range correlations^{18–20}, but an unequivocal experiment ruling out other explanations, such as the observation of a Josephson supercurrent through a fully spin-polarized ferromagnet, has not been reported. In our experiment, we observe such a supercurrent

which prevails over very long length scales ~ 1 μm , that is much longer than expected for singlet correlations, and which characteristically depends on the orientation of the magnetization in the ferromagnet. Therefore, we attribute this long-range supercurrent to superconducting triplet correlations.

In conventional Josephson junctions, the supercurrent I_s is given by $I_c \sin(\phi_1 - \phi_2)$, meaning that the maximum value, the critical current I_c , is obtained for a phase difference $\phi_1 - \phi_2 = \pi/2$, with ϕ_1 the quantum phase of superconductor 1 and ϕ_2 the quantum phase of superconductor 2. The application of an external magnetic field, $\mathbf{H}$, creates a position dependence of the phases ϕ_1 and ϕ_2 , leading to a characteristic periodic dependence on magnetic induction, $\mathbf{B}$, known as a Fraunhofer pattern² (in analogy to optical diffraction). We have applied a magnetic field in the plane of the CrO₂ film. Besides the conventional Fraunhofer pattern, we observe additional effects due to the finite magnetization, $\mathbf{M}$, of the sample, as shown in Fig. 2c. As evident from the raw data shown in the inset, there are clear signs of hysteresis in the critical current as a function of the magnetic field. Owing to the biaxial symmetry of the magnetic system, the magnetization vector follows a different trajectory along the easy directions of the film (clockwise versus anticlockwise) in the up and down sweep of the magnetic field (Fig. 1d). This results in the magnetization in the upsweep lagging behind with respect to the magnetization in the downsweep, leading to an overall shift in the two sweeps (of the order of 45 mT). After removal of the hysteresis, clear oscillations are visible in the critical current, corresponding to the addition of one flux quantum through the junction area per period of the oscillation of 80 mT. This corresponds to an effective junction length of 240 nm (which is somewhat shorter than the actual length of the junction, 310 nm). Although the periodic dependence is quite analogous to the

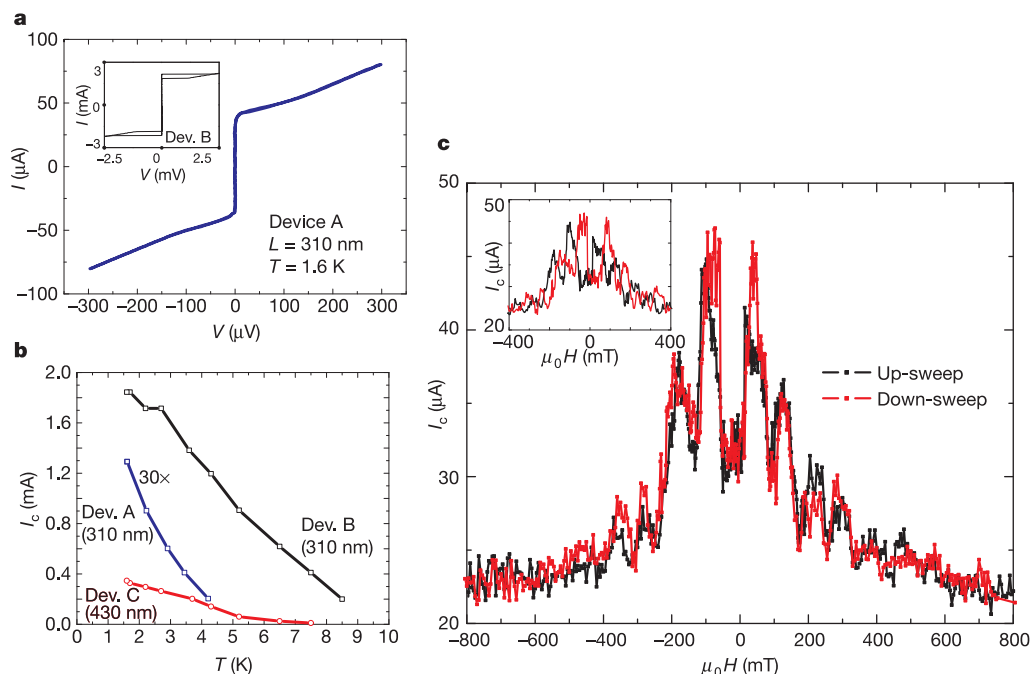


Figure 2 | Observed superconducting transport properties of the superconductor-CrO₂-superconductor system. **a**, Typical current-voltage (I - V) characteristic at temperature $T = 1.6$ K: a zero resistance supercurrent branch is clearly visible (for larger critical currents the current-voltage characteristic is hysteretic, see inset). Similar data have been observed in 10 different samples, some of which had several devices in series. From device to device a spread of critical current of less than 2 orders of magnitude is found. The magnitude of $I_c R_N$ (the product of the critical current, I_c , and the normal state resistance, R_N) is for all junctions smaller than 4 mV (twice the estimated gap size of the NbTiN), and typically 10–300 μV , for nominal junction lengths, L , of 0.3–1 μm . **b**, Critical current as a function of

temperature for three devices. **c**, Critical current as a function of external magnetic field, applied in the plane of the film, for the device used in **a**. The angle ψ between the direction of the current and the field is 90°. The raw data are shown in the inset with a trace for increasing (up-sweep) and one for decreasing (down-sweep) magnetic field strength, demonstrating that we observe hysteresis. In the main figure the measurements for decreasing field strength are shifted by $\mu_0 H = 45$ mT (with the permeability of vacuum, μ_0) with respect to those for increasing field to correct for the hysteresis. The main figure clearly shows oscillations in the critical current with the applied magnetic flux through the junction. For technical reasons no measurements are performed in perpendicular field.

Fraunhofer pattern, there are clear deviations; that is, there is a minimum at the centre rather than a peak. We attribute this effect to the finite sample magnetization, which adds an offset to the flux density, and thus shifts the maximum of the pattern away from $H = 0$.

The biaxial symmetry is also clearly evident from the following experiment. Figure 3 compares measurements obtained by first making a full magnetic field sweep (rotating the magnetization by 360° , 'major loop') and then a sweep that returns halfway ('minor loop'). These data are similar to those obtained in a giant magnetoresistance experiment with not the resistance but the supercurrent being changed by the magnetization. The relative change in I_c of the magneto-switchable Josephson junction is in our case around 30%, and can be further increased by alignment of the junction along an easy axis. Hence, we find that the supercurrent is strongly linked to the magnetization $\mathbf{M}$ of the CrO_2 for a field $\mathbf{H}$ applied in the plane of the film (Fig. 1d). This double dependence on both $\mathbf{H}$ and $\mathbf{M}$ is a unique aspect of these ferromagnetically coupled Josephson junctions.

At present it is difficult to make a quantitative comparison of our measurements to theory. Comparing the temperature dependence of the critical current with the expressions following from the diffusive theory²¹ leads to Thouless energies in the range from 30 to 80 μeV . There is one theoretical paper²² in which triplet correlations are analysed for a geometry resembling our experiment, but only in the limit of one-dimensional ballistic motion, which is not applicable to our samples. Apart from this, a crucial ingredient of theoretical descriptions is the process and the strength of the singlet–triplet conversion. This process is generally believed to be caused by an interplay between the exchange field of the ferromagnet at the superconductor–ferromagnet interface and the presence of a non-homogeneous magnetic field. There exists a large number of proposed mechanisms to create this non-homogeneous field, including, but not limited to, domain walls at the interface, spin–orbit interactions in the superconductor, a rotating magnetization in the ferromagnet close to the interface, and local magnetic impurities. Although we can only speculate which of these mechanisms is

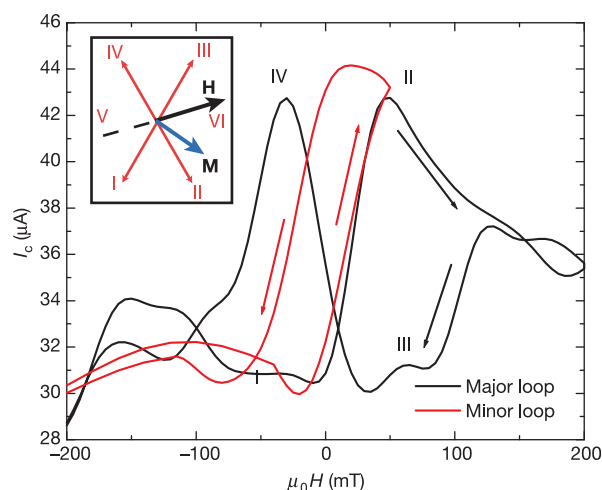


Figure 3 | Control of the critical current by changing the magnetization orientation. The hysteresis loop in the critical current for a sweep from low to high fields ('major loop' from V to VI through I and II and back through III and IV) is different from the hysteresis for a sweep from low to moderately high fields ('minor loop' from V to II and back), with the Roman capitals indicating the axes of the biaxial symmetry (inset). These data illustrate that the CrO_2 behaves as a single domain. The data are for clarity presented as a 5-point average.

effective in our case, the large spread in the critical currents observed in our experiments suggests that the process creating the inhomogeneous field—and thus the process responsible for the singlet–triplet conversion—is poorly defined. This suggests that the formation of the interface plays a crucial role.

Received 22 July; accepted 29 November 2005.

- Shimizu, K. *et al.* Superconductivity in the non-magnetic state of iron under pressure. *Nature* **412**, 316–318 (2001).
- Tinkham, M. *Introduction to Superconductivity* 2nd edn (McGraw-Hill, New York, 1996).
- Bergeret, F. S., Volkov, A. F. & Efetov, K. B. Long-range proximity effects in superconductor-ferromagnet structures. *Phys. Rev. Lett.* **86**, 4096–4099 (2001).
- Bergeret, F. S., Volkov, A. F. & Efetov, K. B. Odd triplet superconductivity and related phenomena in superconductor-ferromagnet structures. *Rev. Mod. Phys.* **77**, 1321–1373 (2005).
- Iosad, N. N. *et al.* Properties of $(\text{Nb}_{0.35}\text{Ti}_{0.15})_x\text{N}_{1-x}$ thin films deposited on silicon wafers at ambient substrate temperature. *J. Appl. Phys.* **88**, 5756–5759 (2000).
- Coey, J. M. D. & Venkatesan, M. Half-metallic ferromagnetism: example of CrO_2 . *J. Appl. Phys.* **91**, 8345–8350 (2002).
- Ryazanov, V. V. Josephson superconductor-ferromagnet-superconductor π -contact as an element of a quantum bit (experiment). *Phys. Usp.* **42**, 825–827 (1999).
- Kontos, T., Aprili, M., Lesueur, J. & Grison, X. Inhomogeneous superconductivity induced in a ferromagnet by proximity effect. *Phys. Rev. Lett.* **86**, 304–307 (2001).
- Korotin, M. A., Anisimov, V. I., Khomskii, D. I. & Sawatzky, G. A. CrO_2 : a self-doped double exchange ferromagnet. *Phys. Rev. Lett.* **80**, 4305–4308 (1998).
- Gupta, A., Li, X. W. & Xiao, G. Magnetic and transport properties of epitaxial and polycrystalline chromium dioxide thin films (invited). *J. Appl. Phys.* **87**, 6073–6078 (2001).
- Soulen, R. J. Jr *et al.* Measuring the spin polarization of a metal with a superconducting point contact. *Science* **282**, 85–88 (1998).
- Parker, J. S., Watts, S. M., Ivanov, P. G. & Xiong, P. Spin polarization of CrO_2 at and across an artificial barrier. *Phys. Rev. Lett.* **88**, 196601 (2002).
- Spinu, L., Srikanth, H., Gupta, A., Li, X. W. & Xiao, G. Probing magnetic anisotropy effects in epitaxial CrO_2 thin films. *Phys. Rev. B* **62**, 8931–8934 (2000).
- Kadigrobov, A., Shekhter, R. I. & Jonson, M. Quantum spin fluctuations as a source of long-range proximity effects in diffusive ferromagnet-superconductor structures. *Europhys. Lett.* **54**, 394–400 (2001).
- Leggett, A. J. A theoretical description of the new phases of liquid ^3He . *Rev. Mod. Phys.* **47**, 331–414 (1975).
- MacKenzie, A. P. & Maeno, Y. The superconductivity of Sr_2RuO_4 and the physics of spin-triplet pairing. *Rev. Mod. Phys.* **75**, 657–712 (2003).
- Saxena, S. S. *et al.* Superconductivity on the border of itinerant-electron ferromagnetism in UGe_2 . *Nature* **406**, 587–592 (2000).
- Giroud, M., Courtois, H., Hasselbach, K., Mailly, D. & Pannetier, B. Superconducting proximity effect in a mesoscopic ferromagnetic wire. *Phys. Rev. B* **58**, R11872–R11875 (1998).
- Petrashov, V. T., Sosnin, I. A., Cox, I., Parsons, A. & Troadec, C. Giant mutual proximity effects in ferromagnetic/superconducting nanostructures. *Phys. Rev. Lett.* **83**, 3281–3284 (1999).
- Peña, V. *et al.* Coupling of superconductors through a half-metallic ferromagnet: Evidence for a long-range proximity effect. *Phys. Rev. B* **69**, 224502 (2004).
- Wilhelm, F. K., Zaikin, A. D. & Schön, G. Supercurrent in a mesoscopic proximity wire. *J. Low-Temp. Phys.* **106**, 305–310 (1997).
- Eschrig, M., Kopu, J., Cuevas, J. C. & Schön, G. Theory of half-metal/superconductor heterostructures. *Phys. Rev. Lett.* **90**, 137003 (2003).
- Li, X. W., Gupta, A. & Xiao, G. Influence of strain on the magnetic properties of epitaxial (100) chromium dioxide (CrO_2) films. *Appl. Phys. Lett.* **75**, 713–715 (1999).

Acknowledgements We thank I. van Dijk, M. G. Flokstra, H. T. Man and S. Russo for stimulating interactions. This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). The work at the University of Alabama was supported by the National Science Foundation as part of an MRSEC grant.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S.K. (r.s.keizer@tnw.tudelft.nl) or T.M.K. (t.m.klapwijk@tnw.tudelft.nl).

LETTERS

Scale invariance and universality of force networks in static granular matter

Srdjan Ostojic¹, Ellák Somfai^{2†} & Bernard Nienhuis¹

Force networks form the skeleton of static granular matter^{1,2}. They are the key factor that determines mechanical properties such as stability³, elasticity^{4,5} and sound transmission^{6,7}, which are important for civil engineering and industrial processing. Previous studies have focused on investigations of the global structure of external forces^{8–11} (the boundary condition) and on the probability distribution of individual contact forces^{4,12}. So far, however, precise knowledge of the disordered spatial structure of the force network has remained elusive. Here we report that molecular dynamics simulations of realistic granular packings reveal scale invariance of clusters of particles interacting by means of relatively strong forces. Despite visual variation, force networks for various values of the confining pressure and other parameters have identical scaling exponents and scaling function, thereby determining a universality class. Unexpectedly, the flat ensemble of force configurations^{13–15} (a simple generalization of equilibrium statistical mechanics) belongs to this universality class, whereas some widely studied simplified models^{16–18} do not. This implies that the elasticity of the grains and their geometrical disorder do not affect the universal mechanical properties.

Sand dunes, piles of apples in a supermarket and coffee beans in a jar are all examples of static granular matter. In these examples, the particles (sand grains, apples, coffee beans) interact with their neighbours with repulsive contact forces forming a network. Owing to disorder (even in a regular pile, each apple is slightly different), the force network is inhomogeneous in space^{19,20}. The simplest characterization of the force network is the probability distribution of individual contact forces^{21–23}: $P(F)$. In typical granular packings, $P(F)$ has an exponential tail at large F , and a plateau at small F . Although $P(F)$ is an important and well-studied quantity, it tells us nothing about the spatial structure of the force network, which is so striking visually (Fig. 1a). As a large contact force on one side of a particle is typically balanced by another large force on the opposite

side, large forces tend to align in filamentary structures called ‘force chains’, although they are not completely linear. The force chains are stabilized by weaker side contacts and form a disordered branching network resembling a fractal. Typical fractals are invariant under a change of length scale. A quantitative analysis of that scale invariance is, however, difficult, in part because there is no sharp distinction between the force chains and the background contacts.

Here we show that discriminating between forces of different magnitudes provides a quantitative characterization of force chains. By an analogy with equilibrium critical phenomena, such a description naturally captures scale invariance: a set of scaling exponents (fractal dimensions) and a scaling function are extracted from a given ensemble of force networks. We applied this approach to two-dimensional granular packings under isotropic pressure generated by molecular dynamics simulations. Although the appearance of the force network varies significantly with pressure and polydispersity, we found that the scaling exponents and the scaling function are independent of these and other parameters, and thus capture universal properties of force networks. We also show that a theoretical model based on the extension of equilibrium statistical mechanics proposed by Edwards faithfully represents this universality class, whereas some other simple models do not.

The force network in a granular packing can be represented as a set of bonds connecting grains in contact. Each bond carries a scalar given by the magnitude of the repulsive force between the grains (Fig. 1). To quantify the (visually obvious) force chains in such a network, we choose a threshold f and consider only chains formed by bonds carrying forces larger than f . Rather than select a fixed arbitrary value of f , we study such structures at different scales by varying the threshold. For small values most of the grains remain connected, but as the threshold is increased the packing breaks up into disconnected clusters (Fig. 2). The extent of each force chain can then be characterized by the size of the corresponding cluster, that is, the number of mutually connected bonds.

The force network varies totally between packings created under

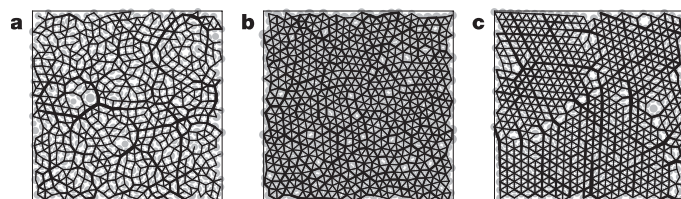


Figure 1 | Dependence of the force network on pressure and polydispersity. Grains are represented as grey disks and forces as bonds. The thickness of each bond is proportional to the magnitude of the force. Shown is packing under pressure $p = 10^{-4}$ (in rescaled units) and polydispersity $d = 20\%$ (a), $p = 10^{-1}$ and $d = 20\%$ (b), and $p = 10^{-2}$ and $d = 5\%$ (c).

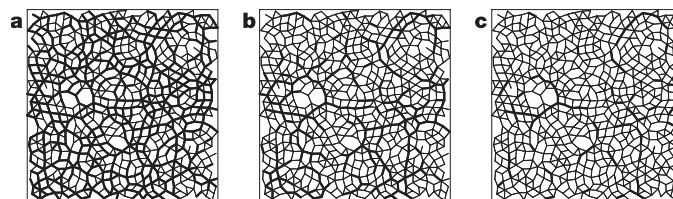


Figure 2 | Force chains at different scales. The packing of Fig. 1a is repeated, with forces larger than a threshold f shown as bold lines. The threshold values are $f = 0.5f_c$ (a), $f = f_c$ (b) and $f = 1.5f_c$ (c), where f_c is the critical threshold.

¹Institute for Theoretical Physics, Universiteit van Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands. ²Universiteit Leiden, Instituut-Lorentz, PO Box 9506, 2300 RA Leiden, The Netherlands. [†]Present address: The Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

the same external conditions, so that a statistical approach is needed. Following the above procedure, force chains in an ensemble of packings can be described by the probability $P(s, f)$ that, at a threshold f , a given random grain is of size s . Similar statistical descriptions in terms of clusters are common in the theory of equilibrium phase transitions²⁴, where the threshold f takes the role of temperature. As in a percolation model²⁵, a phase transition occurs at the critical threshold f_c , above which no cluster connects the opposite boundaries of the system. Around this value, the system shows scale invariance; in particular, the probability distribution of cluster sizes can be expressed as $P(s, f) \approx s^{-\tau} \rho(s/(f - f_c)^\sigma)$. This scale invariance is ‘universal’: the scaling exponents τ and σ , as well as the scaling function ρ , are determined only by the global symmetries of the system rather than the microscopic details of the interactions. Equilibrium critical phenomena can therefore be classified in a discrete number of universality classes according to the values of the exponents.

Such an analogy with equilibrium critical phenomena suggests the existence of scale invariance around the threshold f_c , below which one of the clusters spans the whole packing. The associated scaling exponents and scaling function would provide a new characterization of the spatial organization of forces in granular matter. To study the existence of scale invariance and universality, we examine the behaviour of the n th moment m_n of $P(s, f)$ as function of the threshold f and the system size N (number of contacts in the packing). We look for scaling as function of the system size, which in case of equilibrium critical phenomena takes the form

$$m_n(f, N) \approx N^{\phi_n} M_n([f - f_c]N^{1/2\nu}) \quad (1)$$

where the scaling function M_n is related by integration to ρ , and the scaling exponents are given by $\phi_n = (n + 1 - \tau)/(\tau - 1)$ and $\nu = (\tau - 1)/2\sigma$. Here we present a study for $n = 2$ because it is the lowest moment to diverge, but the results hold for higher moments.

We have carried out a scaling analysis on packings created by molecular dynamics simulations (also called discrete elements methods) of a system of polydisperse spheres in a periodic two-dimensional cell subject to an isotropic pressure. The particles are deformable and interact by means of nonlinear Hertz–Mindlin forces (see Methods). We first consider strongly polydisperse frictionless packings at low pressure, leading to strongly disordered force networks with clearly visible force chains (Fig. 1a). For different system sizes, we determined the force chain clusters at a large number of threshold values and thus computed $m_2(f, N)$, the second moment of clusters sizes (leaving out the largest cluster in each sample). If m_2 scales according to equation (1), then plotting $N^{-\phi} m_2$ as function of $([f - f_c]N^{1/2\nu})$ should lead to a collapse of data for different system sizes on a single curve. Varying ϕ , ν and f_c , a good collapse is indeed obtained for $\phi = 0.89 \pm 0.01$ and $\nu = 1.6 \pm 0.1$, clearly confirming scale invariance.

To test the dependence on various parameters, we repeated the procedure for packings created at different pressures, polydispersities, coefficients of friction and force laws. As the pressure is increased, the grains deform and the number of contacts per particle rises (Fig. 1b). The force network becomes increasingly uniform in appearance, and the distribution of force magnitudes becomes narrower (see Supplementary Information). Computing $m_2(f)$ for values of pressure spanning three orders of magnitude, we nevertheless find that the optimal data collapse is obtained in the same range, $\phi = 0.89 \pm 0.01$ and $\nu = 1.6 \pm 0.1$, independently of the pressure. Decreasing the polydispersity leads to crystallization of the grains (Fig. 1c); however, such ordering of the contact network also leaves the values of the scaling exponents unchanged. Moreover, although friction leads to smaller coordination number in the packing, it does not influence the scaling properties. In this case the forces are not normal to the grains, but tangential forces are typically much smaller than f_c (see Supplementary Information), so that they do not affect the scaling of cluster sizes. Finally, particles

interacting with harmonic forces instead of nonlinear Hertz forces lead once again to the same exponents.

The independence of the exponents on pressure, polydispersity, friction and force law is clear evidence of universality. As expected, however, the value of the critical threshold is not universal: it varies between 1.3 and 1.6 times the average normal force and depends on all the parameters. Moreover, the width and height of the scaling function vary with the parameters: the width decreases with increasing pressure, whereas the height increases with increasing polydispersity. A linear rescaling of both axes nevertheless leads to a single collapse for all considered values of parameters (Fig. 3): similarly to equilibrium critical phenomena, the full scaling function seems to be universal.

As the scaling exponents and the scaling function are universal, they can be obtained from simplified models in the appropriate universality class. Years ago, Edwards proposed to generalize the micro-canonical principle of thermodynamics to jammed systems such as granular matter¹³. The key idea is to ignore history and to treat as equally likely all stable configurations of grains. The approach was found to be successful in slowly flowing granular matter¹⁴, but so far a quantitative confirmation for static granular matter has been lacking.

A simple extension of this micro-canonical ensemble to a packing of grains in a fixed geometry has been named the ‘snooker packing’¹⁵. It consists of a hexagonal arrangement of rigid, frictionless and monodisperse spheres confined within a triangular domain, with the same pressure applied to all sides of the packing. In such a system, force balance alone does not completely determine the forces. Here it is natural to generalize Edwards’ approach by treating as equally likely all arrangements of repulsive forces in balance on each grain, while keeping fixed the geometry of contacts. Using Monte Carlo simulations (see Methods), we generated force networks on the snooker packing with uniform probability, and examined the scaling of the second moment of cluster size distribution for different system sizes. The optimal collapse is obtained for values of scaling exponents

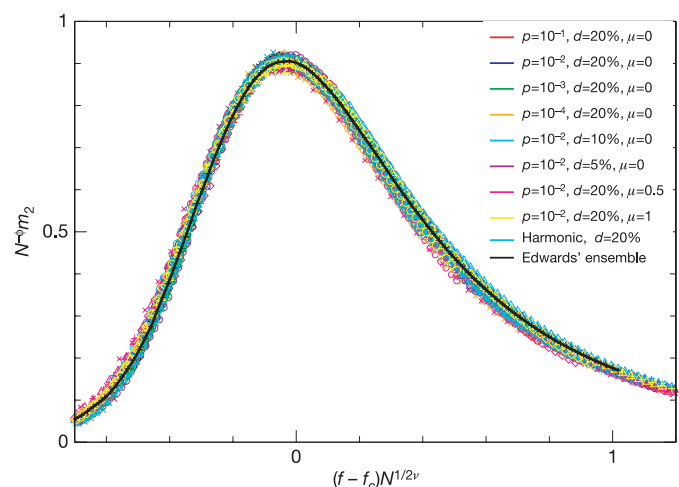


Figure 3 | The universal scaling function. The mean square of the cluster sizes, omitting the largest cluster, m_2 , is rescaled as $BN^{-\phi} m_2$ and plotted as function of the rescaled threshold $A([f - f_c]N^{1/2\nu})$. The graph shows the collapse on a single curve of 50 data sets corresponding to various packing parameters and system sizes. A different colour is associated with each of six sets of parameters (pressure p , polydispersity d , coefficient of friction μ) used in molecular dynamics simulations. For each set of parameters, five system sizes (increasing by factors of two) are represented by symbols: circle, diamond, triangle, cross and plus. The collapse was obtained for $\phi = 0.89 \pm 0.01$ and $\nu = 1.6 \pm 0.1$ for all data sets, whereas the values of f_c , A and B depend on the parameters (but not on the system size). In black are shown Monte Carlo results of the Edwards’ ensemble on the ‘snooker packing’, which gives the same values of ϕ and ν .

within error bars equal to those found from molecular dynamics simulations. Moreover, the full scaling function appears to be identical to the one found in molecular dynamics simulations (Fig. 3).

Some other simple models of granular matter lead to different scaling exponents. For example, the extensively studied q model^{16–18}, which implements balance between vertical forces only, leads to $\phi = 0.69 \pm 0.01$ and $\nu = 3.1 \pm 0.1$. Moreover, the universality class depends on whether a top-to-bottom propagation of forces is assumed (S.O. and B.N., manuscript in preparation). The agreement between the Edwards ensemble and the results of molecular dynamics simulations is thus rather remarkable. In particular, it shows that the elasticity of the grains and the details of the force that it generates are irrelevant for the shape of the scaling function. Moreover, the geometrical randomness of the grain arrangements does not have any bearing on the universal mechanical properties. We expect that the key ingredients are the vectorial balance of forces on each grain and the isotropy of the confining pressure: static packings under shear may lead to different scaling exponents. We conjecture that in three dimensions the exponents and scaling function are equally universal, although different from their two-dimensional equivalent.

Our results establish an unexpected connection between static granular matter and equilibrium critical phenomena. The observed scale invariance is in all aspects comparable to equilibrium criticality and defines a novel universality class of isotropic force networks. This universality class is for example distinct from that of percolation, where the same analysis can be applied. This result implies the existence of long-range correlations between forces, a characteristic of structures such as force chains. We expect force networks in other jammed systems, such as foams and emulsions, to belong to the same universality class. We hope that progress in the measurement of intergrain and interbubble forces^{12,26} will lead to an experimental evaluation of the scaling exponents.

METHODS

Molecular dynamics. We use molecular dynamics simulations (also known as a discrete element method²⁷ in engineering) to create static granular packings that mimic real granular matter. Starting from a dilute gas phase, we solve Newton's equations where the particles interact with a frictionless repulsive contact force, called the hertzian interaction²⁸. The centres of the particles are restricted to a two-dimensional plane; however, the hertzian repulsive force is that of three-dimensional elastic spheres: $F = (2/3)E/(1 - \nu^2)R^{1/2}n^{3/2}$, where E and ν are the Young's modulus and Poisson ratio of the particle's material, and $R = R_1 R_2 / (R_1 + R_2)$. The overlap $n = R_1 + R_2 - r_{12}$ is calculated from the particle's radii R_1 and R_2 and the distance between their centres r_{12} . In part of the analysis, we used the frictional Hertz–Mindlin²⁸ force law. During the evolution, we shrink the volume of the periodic box with rate controlled by a feedback loop to achieve a target pressure. Because we include some dissipation in the particle interactions, eventually all motion stops and a static granular packing is obtained in mechanical equilibrium. For each set of parameters, we use 100 independent packings of 10,000 particles in two dimensions, with polydispersity ranging from 5 to 20% (maximum deviation from the mean, using flat distribution of radii). By extracting recursively subsystems with half the number of grains, in total five system sizes were obtained. The pressures are expressed in units of the Young's modulus of the particle's material. For more details see ref. 7.

Monte Carlo. The Edwards ensemble for a granular heap is a uniform probability measure on the forces subject to equations for the balance of each particle and inequalities for the positivity of each contact force. In an appropriately reduced space of forces, this represents a constant nonzero probability density inside a convex polyhedron. We sample this distribution by the following stochastic process. Starting with a point inside the polyhedron, we select a line in a random direction (from a discrete set of directions). In this direction we calculate the intersections with the polyhedron. A new point is chosen on the line with uniform probability between the intersection points. Averages calculated from such a process converge to unweighted averages over the whole polyhedron. For more details see ref. 29.

Received 26 August; accepted 22 December 2005.

- Garcia-Rojo, R., Herrmann, H. J. & McNamara, S. (eds) *Powders and Grains* 2005 (Balkema, Leiden, 2005).
- Jaeger, H. M., Nagel, S. R. & Behringer, R. P. Granular solids, liquids and gases. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
- Daerr, A. & Douady, S. Two types of avalanche behaviour in granular media. *Nature* **399**, 241–243 (1999).
- Makse, H. A., Gland, N., Johnson, D. L. & Schwartz, L. M. Why effective medium theory fails in granular materials. *Phys. Rev. Lett.* **83**, 5070–5073 (1999).
- Goldenberg, C. & Goldhirsch, I. Friction enhances elasticity in granular solids. *Nature* **435**, 188–191 (2005).
- Jia, X., Caroli, C. & Velicky, B. Ultrasound propagation in externally stressed granular media. *Phys. Rev. Lett.* **82**, 1863–1866 (1999).
- Somfai, E., Roux, J.-N., Snoeijer, J. H., van Hecke, M. & van Saarloos, W. Wave propagation in confined granular systems. *Phys. Rev. E* **72**, 021301 (2005).
- Vanel, L., Howell, D., Clark, D., Behringer, R. P. & Clement, E. Memories in sand: Experimental tests of construction history on stress distributions under sandpiles. *Phys. Rev. E* **60**, R5040–R5043 (1999).
- Wittmer, J. P., Claudin, P., Cates, M. E. & Bouchaud, J.-P. An explanation for the central stress minimum in sand piles. *Nature* **382**, 336–339 (1996).
- Geng, J. et al. Footprints in sand: the response of a granular material to local perturbations. *Phys. Rev. Lett.* **87**, 035506 (2001).
- Reydellet, G. & Clement, E. Green's function probe of a static granular piling. *Phys. Rev. Lett.* **86**, 3308–3311 (2001).
- Majumdar, T. S. & Behringer, R. P. Contact force measurements and stress-induced anisotropy in granular materials. *Nature* **435**, 1079–1082 (2005).
- Edwards, S. F. & Oakeshott, R. Theory of powders. *Physica A* **157**, 1080–1090 (1989).
- Makse, H. A. & Kurchan, J. Testing the thermodynamics approach to granular matter with a numerical model of a decisive experiment. *Nature* **415**, 614–617 (2002).
- Snoeijer, J. H., Vlucht, T. J. H., van Hecke, M. & van Saarloos, W. Force network ensemble: a new approach to static granular matter. *Phys. Rev. Lett.* **92**, 054302 (2004).
- Liu, C.-h. et al. Force fluctuations in bead packs. *Science* **269**, 513–515 (1995).
- Coppersmith, S. N., Liu, C.-h., Majumdar, S., Narayan, O. & Witten, T. A. Model for force fluctuations in bead packs. *Phys. Rev. E* **53**, 4673–4685 (1996).
- da Silva, M. & Rajchenbach, J. Stress transmission through a model system of cohesionless elastic grains. *Nature* **406**, 708–710 (2000).
- Radjai, F., Jean, M., Moreau, J.-J. & Roux, S. Force distributions in dense two-dimensional granular systems. *Phys. Rev. Lett.* **77**, 274–277 (1996).
- Makse, H. A., Johnson, D. L. & Schwartz, L. M. Packing of compressible granular materials. *Phys. Rev. Lett.* **84**, 4160–4163 (2000).
- Mueth, D. M., Jaeger, H. M. & Nagel, S. R. Force distribution in a granular medium. *Phys. Rev. E* **57**, 3164–3169 (1998).
- Blair, D. L., Mueggenburg, N. W., Marshall, A. H., Jaeger, H. M. & Nagel, S. R. Force distributions in 3D granular assemblies: effects of packing order and inter-particle friction. *Phys. Rev. E* **63**, 041304 (2001).
- Erikson, J. M., Mueggenburg, N. W., Jaeger, H. M. & Nagel, S. R. Force distributions in three-dimensional compressible granular packs. *Phys. Rev. E* **66**, 040301 (2002).
- Fortuin, C. M. & Kasteleyn, P. W. On the random cluster model. *Physica* **57**, 536–564 (1971).
- Stauffer, D. & Aharony, A. *Introduction to Percolation Theory* (Taylor & Francis, London, 1991).
- Bruijic, J. et al. 3D bulk measurements of the force distribution in a compressed emulsion system. *Faraday Disc.* **123**, 207–220 (2003).
- Cundall, P. A. & Strack, O. D. L. A discrete numerical model for granular assemblies. *Géotechnique* **29**, 47–65 (1979).
- Johnson, K. L. *Contact mechanics* (Cambridge Univ. Press, Cambridge, UK, 1985).
- Ostojic, S. & Panja, D. Response of a hexagonal granular packing under a localized external force. *Europhys. Lett.* **71**, 70–77 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Behringer, W. Ellenbroek, C. Goldenberg, M. van Hecke, W. van Saarloos and K. Shundyak for discussions. S.O. is financially supported by the Dutch research organization FOM (Fundamenteel Onderzoek der Materie). E.S. is supported by the PHYNECS training network of the European Commission.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.N. (nienhuis@science.uva.nl).

The distillation and volatility of ionic liquids

Martyn J. Earle¹, José M.S.S. Esperança², Manuela A. Gilea¹, José N. Canongia Lopes², Luís P.N. Rebelo², Joseph W. Magee³, Kenneth R. Seddon¹ & Jason A. Widegren³

It is widely believed that a defining characteristic of ionic liquids (or low-temperature molten salts) is that they exert no measurable vapour pressure, and hence cannot be distilled^{1,2}. Here we demonstrate that this is unfounded, and that many ionic liquids can be distilled at low pressure without decomposition. Ionic liquids represent matter solely composed of ions, and so are perceived as non-volatile substances. During the last decade, interest in the field of ionic liquids has burgeoned³, producing a wealth of intellectual and technological challenges and opportunities for the production of new chemical and extractive processes^{4–6}, fuel cells and batteries⁷, and new composite materials^{8,9}. Much of this potential is underpinned by their presumed involatility. This characteristic, however, can severely restrict the attainability of high purity levels for ionic liquids (when they contain poorly volatile components) in recycling schemes, as well as excluding their use in gas-phase processes. We anticipate that our demonstration that some selected families of commonly used aprotic ionic liquids can be distilled at 200–300 °C and low pressure, with concomitant recovery of significant amounts of pure substance, will permit these currently excluded applications to be realized.

The belief that ionic liquids are not volatile can probably be traced back to a report on first-generation ionic liquids¹⁰ by Øye and co-workers¹¹, that the Franklin acidic [C₂mim]Cl – AlCl₃ system (in which [C₂mim]⁺ is 1-ethyl-3-methylimidazolium) does not exhibit a measurable vapour pressure, even at raised temperatures, whereas the analogous NaCl–AlCl₃ system exhibited a significant vapour pressure of Al₂Cl₆ (refs 11, 12). However, for [C₂mim]Cl – AlCl₃ systems with a greater than 2:1 excess of AlCl₃, a detectable vapour

pressure of Al₂Cl₆ was observed above 191 °C. Again, this is not a vapour of ions, and is generated by dissociating the ionic liquid¹³. With the discovery of second-generation and third-generation ionic liquids, and the explosion of newly discovered systems, this lack of volatility seems to have been assumed rather than tested.

Of course, it is known that ionic liquids can be thermally decomposed, leading to transalkylation (alkyl scrambling) in the condensate. For example, [C₂mim]Cl thermally decomposes when heated to 190 °C in vacuum: the volatile decomposition products (1-methylimidazole, 1-ethylimidazole, chloromethane, chloroethane, ethene and hydrogen chloride) can be recondensed to create a new ternary ionic liquid system: [C₁mim]Cl – [C₂mim]Cl – [C₂eim]Cl, in which [C₁mim]⁺ is 1,3-dimethylimidazolium and [C₂eim]⁺ is 1,3-diethylimidazolium¹⁴. Similarly, ionic liquids with protonated cations, [BH]X (for example, [Hmim]Cl), will dissociate on heating to produce the molecular base, B, and the molecular acid, HX, which can be recondensed to regenerate [BH]X^{15,16}. However, in neither scenario did the ionic liquid exert a vapour pressure—the vapour consisted of uncharged molecular components, not the ionic liquid itself.

We show here that a range of pure, aprotic, ionic liquids can be vaporized under vacuum at 200–300 °C and then recondensed at lower temperatures. Some selected families of ionic liquids show no signs of degradation either in the distillate or the residue. We also demonstrate significant enrichment in distillations of mixtures of ionic liquids. Multiple-ion cluster transfer into the gas phase is most probably the underlying mechanism by which these compounds first evaporate, then maintain their stability in the low-density phase, and, finally, recondense as a pure ionic liquid at lower temperatures¹⁷.

Two types of apparatus were used for these experiments: a Kugelrohr apparatus (Fig. 1) and a sublimation apparatus (see Supplementary Information). Survey experiments were performed at 300 °C and 0.1 to 5 mbar in the Kugelrohr apparatus (1 mbar = 100 Pa). These temperatures and pressures are below the predicted boiling points¹⁸ of the ionic liquids, but it was possible to vaporize the ionic liquids and then condense them on colder parts of the Kugelrohr apparatus. Typically, times for the distillation were 4 to 6 h. A photograph of the apparatus is given in Fig. 1. In this apparatus, there is a possibility of splashing as a result of bubbles bursting in the ionic liquid, and to prevent this from affecting the

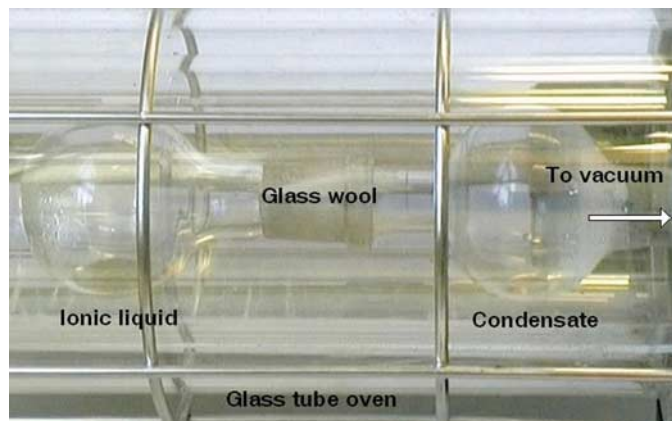


Figure 1 | Labelled photograph of the Kugelrohr oven and distillation apparatus. The central glassware rapidly rotates.

Table 1 | Distillation rates at 300 °C and 0.1 mbar

Second-generation ionic liquid	Distillation rate (g h ⁻¹)
[C ₂ mim][NTf ₂]	0.120
[C ₁₀ mim][NTf ₂]	0.070
[C ₁₆ mim][NTf ₂]	0.024

Results obtained in the Kugelrohr apparatus.

¹The QUILL Centre, The Queen's University of Belfast, Stranmillis Road, Belfast BT9 5AG, UK. ²Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. República, Apartado 127, 2780-901 Oeiras, Portugal. ³Physical and Chemical Properties Division, National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305, USA.

results glass wool was placed between the oven flask and the collection flask.

To demonstrate the distillation of pure ionic liquids, 1.0 g samples of $[\text{C}_2\text{mim}][\text{NTf}_2]$ and $[\text{C}_{10}\text{mim}][\text{NTf}_2]$ (in which $[\text{C}_{10}\text{mim}]^+$ is 1-decyl-3-methylimidazolium, and $[\text{NTf}_2]^-$ is bis{(trifluoromethyl)sulphonyl}amide, also known as bistriflamide, $[\text{N}(\text{SO}_2\text{CF}_3)_2]^-$) were heated for five hours at 300 °C and 0.1 mbar. This process was filmed and two time-lapse (speeded up 100 and 150 times) video files are available in the Supplementary Information to demonstrate the process in action. $[\text{C}_{16}\text{mim}][\text{NTf}_2]$ (in which $[\text{C}_{16}\text{mim}]^+$ is 1-hexadecyl-3-methylimidazolium) was treated similarly. The condensate was collected and analysed by ^1H , ^{19}F and ^{13}C nuclear magnetic resonance (NMR) spectroscopy. These three ionic liquids are stable to prolonged heating at the distillation temperature and could be distilled, although the distillation rate varied according to the chemical structure (Table 1). The NMR spectra of these three compounds showed that no detectable decomposition had occurred during the distillations.

Distillations of a range of other ionic liquids were performed at 300 °C in the Kugelrohr apparatus (Table 2). Generally, the bistriflamide ionic liquids distilled without significant decomposition, and only salts containing the larger tetraalkylammonium ($[\text{N}(\text{C}_{wxyz})^+]$), tetraalkylphosphonium ($[\text{PC}_{wxyz}]^+$), or cholinium ($[\text{C}_{xyz}\text{N}(\text{C}_2\text{H}_4\text{OH})^+]$) cations showed signs of decomposition. For the triflate ($[\text{OTf}]^-$) salts, only $[\text{C}_2\text{dbu}][\text{OTf}]$ (in which $[\text{C}_2\text{dbu}]^+$ is 1-alkyl-1,8-diazabicyclo[5.4.0]undec-7-enium) distilled without significant decomposition. The vapour pressures for triflate salts seemed lower than for similar bistriflamide salts, as shown by their low distillation rates ($<0.01 \text{ g h}^{-1}$). Examples of tosylate ($[\text{OTs}]^-$, $[\text{Me}-4-\text{C}_6\text{H}_4\text{SO}_3]^-$) and hexafluorophosphate ($[\text{PF}_6]^-$) ionic liquids distilled very slowly, with little decomposition (the $[\text{PF}_6]^-$ salt must be free of acidic or basic impurities for a clean distillation). Ionic liquids with other anions (for example, halides, sulphates, or carboxylates) decomposed on distillation. The principal mechanism of decomposition was by dealkylation or transalkylation of the cation¹⁴. This is aided by a nucleophilic anion. The triflate anion has low nucleophilicity, and hence forms relatively thermally stable ionic liquids. The bistriflamide anion has an extremely low nucleophilicity, and hence forms very thermally stable ionic liquids; this also leads to a very low probability of transfer of a proton from the cation to the anion.

To probe the effects of temperature and pressure on the distillation of ionic liquids, three experiments were performed with $[\text{C}_6\text{mim}][\text{NTf}_2]$ in a small glass sublimation apparatus (Table 3). No precautions for splashing were needed in this apparatus because it

is mechanically still and bubble formation was not observed at the temperatures and pressures of these experiments. For each experiment, $[\text{C}_6\text{mim}][\text{NTf}_2]$ was heated from room temperature to 200 or 250 °C, then held at the maximum temperature for 1–2 h. The temperature at which condensation was first observed on the cold finger condenser (or on the walls of the apparatus) was noted. At the end of the experiment, the relative rate of distillation was determined from the amount of $[\text{C}_6\text{mim}][\text{NTf}_2]$ that had condensed on the cold finger. Additionally, the purities of the distillate and of the undistilled residue were checked by ^1H and ^{19}F NMR spectroscopy. The distillation at atmospheric pressure (in dry air) and 250 °C failed in that the distillation was very slow, and much of what ‘distilled’ consisted of decomposition products. On the other hand, the distillation at 200 °C and ≤ 0.001 mbar proceeded much faster, despite being 50 °C lower in temperature, and subsequent NMR analysis showed that the distillate was pure $[\text{C}_6\text{mim}][\text{NTf}_2]$. The pressure–temperature effects on the distillations are intimately related to the vapour pressure of the ionic liquids. A manuscript reporting the determination of the vapour pressure of $[\text{C}_{12}\text{mim}][\text{NTf}_2]$ by a direct static method is under preparation, and a recent paper¹⁹ describing the use of an indirect (Knudsen) method for vapour pressure determination of $[\text{C}_4\text{mim}][\text{NTf}_2]$ has also appeared.

Thus, reduced pressure allowed the distillation to occur more rapidly, at a lower temperature, and without observable decomposition. The unrecognized importance of low pressure to the success of such distillations may have contributed to the myth that ionic liquids cannot be distilled—for example, differential scanning calorimetry (DSC) and thermal gravimetric analysis (TGA) studies (which register the thermal decomposition of ionic liquids) have all been reported at atmospheric pressures.

Thus we have presented experimental data to demonstrate that several ionic liquids can be vaporized and recondensed without significant decomposition. We now consider the mechanism of mass transfer. Below, we rule out the three conceivable alternative mechanisms of mass transfer that could explain our results without involving the vaporization of the ionic liquids as ionic species.

We have eliminated the possibility of physical transfer (either by ‘spraying’, ‘splashing’ or ‘creeping’) by investigating the distillation of equimolar binary mixtures of ionic liquids (see Table 4). The observed difference in the compositions of the distillate and residue in such experiments cannot be explained by a physical transfer mechanism. For all of these distillations, the composition of the distillate is enriched in one component (or in two ions for the case in which all four ions in the mixture are different) compared to the residue, which is the hallmark of separative distillation.

It could also be proposed that the ionic liquids were not being transferred into the gas phase as ionic species, but were being volatilized as neutral molecular species by a proton transfer mechanism (for example, the BASIL process²⁰ involves protonating 1-methylimidazole to an ionic liquid, $[\text{Hmim}]\text{Cl}$, which dissociates on heating). Examination of Table 2 reveals that ionic liquids derived from tetraalkylammonium cations, *N,N*-pyrrolidinium cations, and 1-alkyl-1,8-diazabicyclo[5.4.0]undec-7-enium cations can be distilled without significant decomposition. For these cation types, there is no feasible mechanism to transfer a proton to the anion; in other words, the only possible mechanism for volatilization of these

Table 2 | Effects of distillation at 300 °C in the Kugelrohr apparatus

Ionic liquid	Conditions	Fraction of residue decomposed (%)	Fraction of distillate decomposed (%)
$[\text{C}_4\text{mim}][\text{NTf}_2]$	8 h, 6 mbar	2	2
$[\text{C}_4\text{dmim}][\text{NTf}_2]$	4 h, 6 mbar	1	1
$[\text{C}_8\text{dbu}][\text{NTf}_2]$	4 h, 6 mbar	3	3
$[\text{N}(\text{C}_{2226})][\text{NTf}_2]$	4 h, 6 mbar	1	50
$[\text{N}(\text{C}_{2222})][\text{NTf}_2]$	4 h, 6 mbar	1	2
$[\text{C}_4\text{mpyrr}][\text{NTf}_2]$	4 h, 6 mbar	1	1
$[\text{PC}_{66614}][\text{NTf}_2]$	4 h, 6 mbar	1	65
$[\text{PC}_{44416}][\text{NTf}_2]$	4 h, 6 mbar	1	60
$[\text{C}_{611}\text{N}(\text{C}_2\text{H}_4\text{OH})][\text{NTf}_2]$	4 h, 6 mbar	15	85
$[\text{C}_2\text{mim}][\text{OTf}]$	4 h, 7 mbar	4	12
$[\text{C}_6\text{mim}][\text{OTf}]$	4 h, 6 mbar	1	50
$[\text{C}_2\text{dbu}][\text{OTf}]$	4 h, 8 mbar	1	1
$[\text{C}_4\text{mim}][\text{PF}_6]$	4 h, 6 mbar	1	1
$[\text{C}_4\text{mim}][\text{FAP}]$	18 h, 8 mbar	99	50
$[\text{P}(\text{s}-\text{C}_4)_3\text{C}_1][\text{OTs}]$	4 h, 6 mbar	5	5
$[\text{C}_2\text{mim}][\text{OSO}_3\text{C}_2\text{H}_5]$	5 h, 7 mbar	99	99
$[\text{PC}_{66614}][\text{decanoate}]$	6 h, 7 mbar	98	95

$[\text{C}_n\text{mpyrr}]^+$ is 1-alkyl-1-methylpyrrolidinium; $[\text{FAP}]^-$ is trifluorotriperfluoroethylphosphate. A fraction of 1% represents the limit of our detection, and is thus an upper bound.

Table 3 | Pressure effects on the distillation of $[\text{C}_6\text{mim}][\text{NTf}_2]$

p (mbar)	t_{max} (°C)	t at which distillation was first observed (°C)	Distillation rate (mg h^{-1})	Purity of the distillate (%)
830	250 ± 5	250 ± 5	<1	~ 50
0.07	200 ± 5	170 ± 5	3	>99
≤ 0.001	200 ± 5	155 ± 5	15	>99

Results obtained in the sublimation apparatus.

Table 4 | Distillation of approximately equimolar mixtures of ionic liquids

Ionic liquid A	Ionic liquid B	Composition ratios of the residue (A:B)	Composition ratios of the distillate (A:B)	Fraction of distillate decomposed (%)
[C ₂ mim][NTf ₂] ^{†‡}	[C ₁₆ mim][NTf ₂]	24:76	76:24	<1
[C ₂ mim][NTf ₂] [§]	[C ₆ mim][NTf ₂]	51:49	59:41	<1
[C ₄ mpyrr][NTf ₂] [‡]	[NC _{2 2 2 2}][NTf ₂]	47:53	53:47	<1
[C ₄ mim][NTf ₂] [§]	[C ₄ mim][PF ₆]	52:48	98:2	<1
[C ₂ mim][OTf] [‡]	[NC _{2 2 2 2}][NTf ₂]	*	*	<1

[†]A mass balance of greater than 99% was observed when 40% of the mixture had distilled. Fractional distillation led to a distillate which was 97:3 (A:B).

[‡]This mixture was distilled in the Kugelrohr apparatus at 300 °C and 0.05 mbar for 6–8 h.

[§]This mixture was distilled in the sublimation apparatus at 190–200 °C and ≤0.001 mbar for 1–3 h.

*The undistilled residue contained 56% [C₂mim]⁺, 44% [NC_{2 2 2 2}]⁺, 47% [OTf][−] and 53% [NTf₂][−]; the distillate contained 30% [C₂mim]⁺, 70% [NC_{2 2 2 2}]⁺, 57% [OTf][−], and 43% [NTf₂][−].

ionic liquids is as intact ions, either alone or aggregated. This is represented schematically in Fig. 2.

In the case of tetraalkylphosphonium-based ionic liquids, the majority decomposed before distillation could occur. However, in the case of tris(*sec*-butyl)methylphosphonium tosylate, distillation without significant decomposition was observed, and again there is no possible mechanism which can be envisaged involving proton transfer. Finally, the possibility of dissociative alkyl transfer was eliminated, as there was no evidence for alkyl scrambling in any experiment (such as when [C₂mim]Cl is distilled)¹⁴.

In the cases of 1-alkyl-3-methylimidazolium salts, the majority distilled without decomposition. However, here a proton transfer mechanism must be considered, as the possibility of transferring the 2-H proton from the cation to the anion, creating a carbene and a free acid, is feasible. Nonetheless, a study of the decomposition of [C₂mim][NTf₂] by pyrolysis mass spectrometry found no evidence for the formation of HNTf₂ at 300 °C (ref. 21). We further note that when we replace the 2-H proton on the cation with a 2-methyl group, the new ionic liquid [C₄dmim][NTf₂] (in which [C₄dmim]⁺ is 1-butyl-2,3-dimethylimidazolium) is both volatile and less decomposed than [C₄mim][NTf₂] (see Table 2). If a proton transfer mechanism were dominant, the 2-methylated ionic liquid would have had its volatility suppressed with a much greater level of decomposition. Indeed, examination of Table 4 reveals that in a mixture of tetraethylammonium bistriflate and 1-ethyl-3-methylimidazolium triflate, a tetraethylammonium ionic liquid

vaporizes preferentially to an imidazolium ionic liquid, making a proton transfer mechanism appear highly improbable.

Thus, if we examine all the reported cases, we can completely discount direct physical transfer of liquids, and also eliminate the possibility of dissociative mechanisms (either alkyl or proton transfer) for all ionic liquids except those based around the 1-alkyl-3-methylimidazolium cation, and even in that case proton transfer must be considered to be highly unlikely. For these latter ionic liquids, we believe that there is no significant contribution from dissociation of the cation to form carbenes, but at this stage it cannot be entirely eliminated.

These distillations, accompanied by the recovery of substantial amounts of commonly employed ionic liquids, corroborate recent theoretical predictions¹⁸ that some selected families of ionic liquids could boil at a temperature sufficiently low to avoid decomposition. These observations lay to rest a paralysing and invalid assumption that has dominated and restricted the field of ionic liquids since its origins, and should open up new ways for exploiting their properties. At the very least, a new method for purifying ionic liquids now exists, and we can imagine other applications (such as isolation of highly soluble products by high-temperature crystallization). However, near ambient temperature the vapour pressure of ionic liquids remains negligible, so we have, in effect, the best of both worlds.

Received 26 August; accepted 21 November 2005.

- Seddon, K. R. Room-temperature ionic liquids—neoteric solvents for clean catalysis. *Kinet. Catal.* **37**, 693–697 (1996).
- Seddon, K. R. in *Molten Salt Forum: Proceedings of 5th International Conference on Molten Salt Chemistry and Technology* (ed. Wendt, H.) 53–62 (Trans Tech Publications, Zürich, 1998).
- ISI Essential Science Indicators Special Topics: *Ionic Liquids* (<http://www.esi-topics.com/ionic-liquids/index.html>) (ISI, 2004).
- Rogers, R. D. & Seddon, K. R. (eds) *Ionic Liquids IIIB: Fundamentals, Progress, Challenges, and Opportunities—Transformations and Processes* (American Chemical Society, Washington DC, 2005).
- Rogers, R. D. & Seddon, K. R. (eds) *Ionic Liquids IIIA: Fundamentals, Progress, Challenges, and Opportunities—Properties and Structure* (American Chemical Society, Washington DC, 2005).
- Atkins, M. P. et al. *Ionic Liquids: A Map for Industrial Innovation: Q001* (January 2004) (QUILL, Belfast, 2004).
- Ohno, H. (ed.) *Electrochemical Aspects of Ionic Liquids* (Wiley-Interscience, Hoboken, 2005).
- Rogers, R. D. & Seddon, K. R. (eds) *Ionic Liquids: Industrial Applications for Green Chemistry* (American Chemical Society, Washington DC, 2002).
- Rogers, R. D., Seddon, K. R. & Volkov, S. (eds) *Green Industrial Applications of Ionic Liquids* (Kluwer, Dordrecht, 2002).
- Wasserscheid, P. & Welton, T. (eds) *Ionic Liquids in Synthesis* (Wiley-VCH, Weinheim, 2003).
- Dymek, C. J. Jr, Hussey, C. L., Wilkes, J. S. & Øye, H. A. in *Joint (Sixth) International Symposium on Molten Salts* (eds Mamantov, G. et al.) 93–104 (The Electrochemical Society, Pennington, New Jersey, 1987).
- Øye, H. A. & Wahnsiedler, W. E. in *The Fourth International Symposium on Molten Salts* (eds Blander, M., Newman, D. S., Saboungi, M. L., Mamantov, G. & Johnson, K.) 58–73 (The Electrochemical Society, Pennington, New Jersey, 1984).
- Øye, H. A., Jagtoyen, M., Oksefjell, T. & Wilkes, J. S. Vapour pressure and thermodynamics of the system 1-methyl-3-ethylimidazolium chloride–aluminium chloride. *Mater. Sci. Forum* **73–75**, 183–190 (1991).

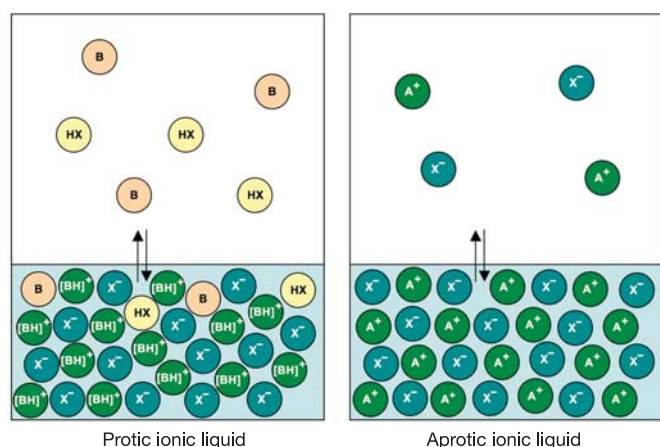


Figure 2 | Schematic representation of the differences between protic and aprotic ionic liquids, in both the liquid and the gaseous phases. For the protic ionic liquids, a dynamic equilibrium exists between the ionic and dissociated forms: $[BH]^+X^-(l) \rightleftharpoons B(l) + HX(l) \rightleftharpoons B(g) + HX(g)$. Green circles represent anions, blue circles represent cations, other coloured circles represent neutral molecules; *l*, liquid phase; *g*, gaseous phase. For the gaseous phase over the aprotic ionic liquid, the representation is purely schematic and has no implication for the actual degree of aggregation.

14. Jeapes, A. J. *et al.* Process for recycling ionic liquids. World Patent WO 01/15175 (2001).
15. Yoshizawa, M., Xu, W. & Angell, C. A. Ionic liquids by proton transfer: Vapor pressure, conductivity, and the relevance of ΔpK_a from aqueous solutions. *J. Am. Chem. Soc.* **125**, 15411–15419 (2003).
16. Kreher, U. P., Rosamilia, A. E., Raston, C. L., Scott, J. L. & Strauss, C. R. Self-associated, “distillable” ionic media. *Molecules* **9**, 387–393 (2004).
17. Abdul-Sada, A. K., Elaiwi, A. E., Greenway, A. M. & Seddon, K. R. Evidence for the clustering of substituted imidazolium salts *via* hydrogen bonding under the conditions of fast atom bombardment mass spectrometry. *Eur. Mass Spectrom.* **3**, 245–247 (1997).
18. Rebelo, L. P. N., Canongia Lopes, J. N., Esperança, J. M. S. S. & Filipe, E. On the critical temperature, normal boiling point, and vapor pressure of ionic liquids. *J. Phys. Chem. B* **109**, 6040–6043 (2005).
19. Paulechka, Y. U., Zaitsau, D. H., Kabo, G. J. & Strechan, A. A. Vapor pressure and thermal stability of ionic liquid 1-butyl-3-methylimidazolium Bis(trifluoromethylsulfonyl)amide. *Thermochim. Acta* **439**, 158–160 (2005).
20. Maase, M. *et al.* Method for the separation of acids from chemical reaction mixtures by means of ionic fluids. World Patent WO 03/062171 (2003).
21. Baranyai, K. J., Deacon, G. B., MacFarlane, D. R., Pringle, J. M. & Scott, J. L. Thermal degradation of ionic liquids at elevated temperatures. *Aust. J. Chem.* **57**, 145–147 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from FC&T, Portugal (to J.M.S.S.E. and L.P.N.R.). The Belfast team wishes to thank the industrial members of QUILL for support and Al Robertson (Cytec) for supplying the phosphonium ionic liquids. For J.W.M. and J.A.W. this work represents an official contribution of the National Institute of Standards and Technology and is not subject to copyright in the USA.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.P.N.R. (luis.rebelo@itqb.unl.pt).

Detection of a direct carbon dioxide effect in continental river runoff records

N. Gedney¹, P. M. Cox², R. A. Betts³, O. Boucher³, C. Huntingford⁴ & P. A. Stott⁵

Continental runoff has increased through the twentieth century^{1,2} despite more intensive human water consumption³. Possible reasons for the increase include: climate change and variability, deforestation, solar dimming⁴, and direct atmospheric carbon dioxide (CO₂) effects on plant transpiration⁵. All of these mechanisms have the potential to affect precipitation and/or evaporation and thereby modify runoff. Here we use a mechanistic land-surface model⁶ and optimal fingerprinting statistical techniques⁷ to attribute observational runoff changes¹ into contributions due to these factors. The model successfully captures the climate-driven inter-annual runoff variability, but twentieth-century climate alone is insufficient to explain the runoff trends. Instead we find that the trends are consistent with a suppression of plant transpiration due to CO₂-induced stomatal closure. This result will affect projections of freshwater availability, and also represents the detection of a direct CO₂ effect on the functioning of the terrestrial biosphere.

On annual and longer timescales, continental river runoff is approximately equal to the difference between land precipitation and evapotranspiration. Changes in river runoff can arise from adjustment to either process. Evapotranspiration is a function of energy and water availability, near-surface atmospheric conditions (air temperature, humidity and wind-speed) and the control of transpiration by plants. Contemporary environmental changes can therefore affect runoff in a number of ways. Climate change and variability modify precipitation patterns and near-surface meteorology. An overall increase in atmospheric aerosol concentration over the past century appears to have reduced the amount of solar radiation reaching the land surface (so-called “solar dimming”⁴), and may have led to reductions in open-pan evaporation⁸. Land-cover changes affect evapotranspiration in a number of ways, including modifying the depth of the soil from which plants can extract water and the available energy by changing the land-surface albedo. Finally, stomatal apertures on plant leaves are observed to close partially under increased CO₂ (ref. 5), suppressing transpiration and providing a mechanism by which CO₂ increase could lead directly to increases in runoff.

Here we analyse observation-based continental river runoff records¹ for evidence of runoff changes from the potential drivers. We adopted the formal detection and attribution techniques developed to isolate the causes of twentieth-century temperature change⁷. A model is used to define the spatial and temporal responses associated with a known forcing factor. These distinct spatio-temporal patterns of response act as ‘fingerprints’ that allow the observed change to be separated into contributions from each factor.

We use the Met Office Surface Exchange Scheme (MOSES II; ref. 6), which is the land-surface scheme in the Hadley Centre climate

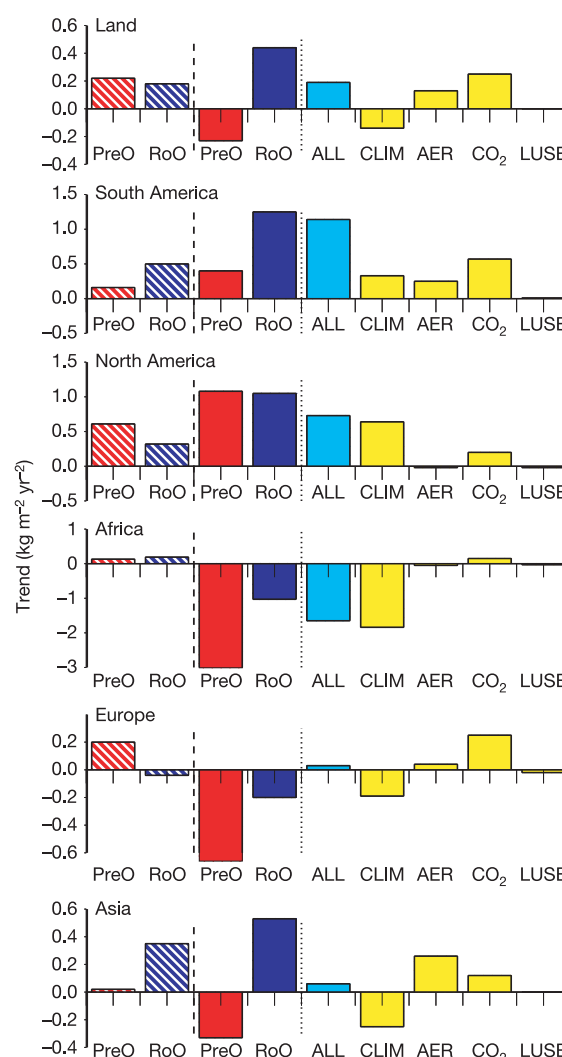


Figure 1 | Trends in continental water budgets. Observed precipitation (PreO; red bars) and runoff (RoO; dark blue bars), and modelled runoff with all mechanisms (ALL; light blue bars) and the individual components (CLIM, AER, CO₂ and LUSE; yellow bars) are shown. ('Land' in the first panel refers to the area-weighted sum of the individual regions.) Striped bars refer to trends over the whole analysis period and solid bars refer to post-1960 trends (see Methods for details).

¹Met Office, Hadley Centre for Climate Prediction and Research (JCHMR), Maclean Building, Wallingford OX10 8BB, UK. ²Centre for Ecology and Hydrology Dorset, Winfrith Technology Centre, Winfrith Newburgh, Dorchester DT2 8ZD, UK. ³Met Office, Hadley Centre for Climate Prediction and Research, Fitzroy Road, Exeter EX1 3PB, UK. ⁴Centre for Ecology and Hydrology Wallingford, Maclean Building, Wallingford OX10 8BB, UK. ⁵Met Office, Hadley Centre for Climate Prediction and Research (Reading Unit), Meteorology Building, University of Reading, Earley Gate, Reading RG6 6BB, UK.

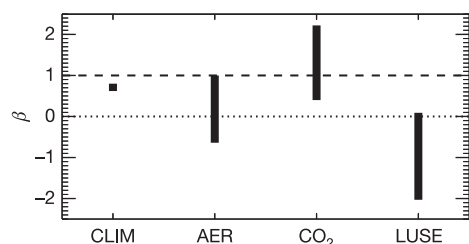


Figure 2 | β scale factors obtained from the optimal fingerprinting technique. The 5 to 95 percentile ranges are shown for CLIM, AER, CO_2 and LUSE.

models. We drive MOSES using a monthly observational data set of the twentieth-century climate⁹. This allows us to assess the extent to which changes in the observed climate can account for changes in runoff. Although here the land-surface scheme is not coupled to a model atmosphere, any feedbacks through the atmospheric boundary layer are implicitly included through changes in the observed near-surface meteorology. MOSES has been extensively validated against data from a number of field sites^{10,11} and performs well globally in its host climate model¹². MOSES is especially appropriate here because it includes a representation of the effect of CO_2 concentration on stomatal conductance¹⁰. We incorporate the effects of aerosols into the forcing data using diagnostics from the transient climate simulation with the HadGEM1 climate model¹³ (see Methods). Changes in land use¹⁴ are used to prescribe related land-cover changes to the MOSES plant types.

To isolate the various effects we carry out five simulations of twentieth-century surface hydrology. We allow all factors likely to affect runoff to vary throughout the fully transient simulation 'ALL' (climate, aerosol concentration, atmospheric CO_2 and land use). In the other four simulations, we allow three of the four factors to vary throughout the twentieth century while fixing the other component to initial conditions (see Methods). The individual effects are found to combine approximately linearly. Hence the impact of each individual component can be calculated by comparing the fixed-component simulation with the ALL simulation (for example, the contribution of CO_2 is given by ALL minus the fixed- CO_2 simulation). CLIM, AER, CO_2 and LUSE refer to the diagnosed contributions of climate, aerosol, atmospheric CO_2 and land use, respectively.

Figure 1 shows the precipitation and modelled and observed runoff trends¹ (see Methods) over the century. The post-1960 observed runoff trends for all regions tend to be more positive (or at least less negative) than the precipitation trends. This suggests that constraints other than precipitation are becoming more important in the latter part of the twentieth century. There is also a clear disparity between the climate-only (CLIM) modelled runoff and the observed runoff trends. With the exception of Europe, which shows a very small reduction in runoff, ALL appears closer to the observations than the climate-only response. The modelled response from climate forcing, aerosols (direct and indirect cloud albedo effects) and direct CO_2 effects appear to be important over some regions. Land use (excluding irrigation) seems to have a small effect on the modelled continental-scale water balance.

We use a standard optimal fingerprinting technique⁷ to see which factors are likely to be driving the long-term changes in continental runoff by comparing the modelled and observed annual anomalies relative to the long-term mean. This technique assumes that we can reproduce the observed anomalies by the linear addition of each individual modelled response x whose amplitude is scaled by a factor β_x . The β factor for each modelled component is obtained by carrying out an ordinary least-squares regression fitted to the observed annual anomalies. This approach assumes that model signals are much less contaminated by noise than are the observed changes⁷, an assumption that we expect to be valid in this case where

observed climatologies are used to force MOSES. If the 5 to 95 percentile range of β is greater than zero then the signal has been 'detected' at the 5% significance level. If this range of β also covers unity, then the modelled component is consistent with the observed response. If, in addition to detection and consistency described above, the signal is not consistent with alternative, physically plausible explanations, then it is said to be 'attributed'¹⁵.

We initially apply the regression over each individual region for the longer time periods given in Table 1 (for example, South America 1903–1994). Because the factors producing only long-term modelled runoff changes all show fairly similar temporal patterns even though their amplitudes are different, we initially carry out the regression analysis with two factors at a time (to avoid over-fitting the data). Only the climate forcing produces considerable inter-annual variability, so we can combine this with each of the other components in separate regressions. The impact of climate change or variation on runoff is detected over all regions at the 5% significance level, and in South America and Asia the model's simulation of runoff due to climate is consistent with that observed (results not shown). The direct CO_2 effect is detected over Africa at the 5% significance level. Aerosols and the CO_2 effect are both consistent with the observed trends over Asia. The probability of $\beta(\text{CO}_2)$ being greater than 0 is higher than for $\beta(\text{AER})$ and $\beta(\text{LUSE})$ for every region except South America, suggestive of the direct CO_2 effect being the most likely global signal.

To isolate these non-CLIM components, we combine the data for each continent so that spatial as well as temporal patterns are used in the fitting process: that is, we carry out a simultaneous best fit across all regions. We now require an estimate of the combined observation and model uncertainty for each region, which we take from the residual statistics calculated in the regional regressions. All the continental regions are considered together, and we carry out a four-component regression for CLIM, AER, CO_2 and LUSE. Only climate and the direct CO_2 effect are detected at the 5% significance level (Fig. 2). (The time series of the fit for each region is shown in Supplementary Fig. S1.) The simulation of runoff change due to the CO_2 increase is consistent with that observed, whereas the model slightly over-estimates the climate effect (since $\beta < 1$) (Fig. 2). We also consider different time series, and because the land-use effect appears to be very small we additionally carry out a three-way regression without land-use changes (not shown). Both these sensitivity studies produce similar β values and result in the same conclusions.

Figure 3 shows the trends attributable to the different factors (that is, with the modelled patterns scaled by β). Most importantly, we find that runoff enhancement due to suppression of transpiration is detected in the observational records (Fig. 3). The records are not consistent with alternative explanations that exclude the effects of increasing CO_2 on plant stomatal conductance. Therefore, we attribute increases in continental runoff to anthropogenic effects via the suppression of transpiration. This is consistent with the fact that there is only limited evidence for anthropogenic effects on global precipitation¹⁶.

Other mechanisms not considered here could also contribute to

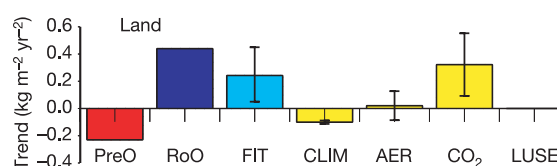


Figure 3 | Attribution of post-1960 overall runoff trend. The first two bars are the observed precipitation (red bar) and runoff (dark blue bar) trends. The remaining bars are the attributed runoff trends for the best fit (FIT; light blue bar) and its individual components (yellow bars). The 5 to 95 percentile ranges are shown.

Table 1 | Trends in observational data

	Time period	Shortwave radiation	Net radiation	Precipitation	Runoff
South America	1903–94	-1.12 ± 0.06 (–0.18)	-0.91 ± 0.03 (–0.12)	0.16 ± 0.16	0.50 ± 0.27
	1960–94	-2.27 ± 0.26 (–0.78)	-1.35 ± 0.10 (–0.19)	0.40 ± 0.78	1.24 ± 1.19
North America	1901–94	-1.64 ± 0.08 (–1.23)	-0.37 ± 0.03 (–0.02)	0.61 ± 0.12	0.32 ± 0.07
	1960–94	-0.01 ± 0.20 (–0.55)	0.41 ± 0.07 (0.00)	1.08 ± 0.45	1.05 ± 0.25
Africa	1901–84	-1.71 ± 0.09 (–0.01)	-1.20 ± 0.05 (–0.04)	0.13 ± 0.13	0.19 ± 0.06
	1960–84	-4.24 ± 0.40 (–1.09)	-2.46 ± 0.14 (–0.42)	-3.01 ± 0.89	-1.02 ± 0.41
Europe	1901–94	-1.54 ± 0.07 (–0.09)	-1.25 ± 0.05 (0.01)	0.20 ± 0.14	-0.04 ± 0.10
	1960–94	-0.18 ± 0.32 (–0.26)	0.04 ± 0.21 (0.03)	-0.66 ± 0.55	-0.20 ± 0.46
Asia	1930–94	-1.97 ± 0.09 (–0.17)	-1.25 ± 0.03 (0.00)	0.09 ± 0.12	0.37 ± 0.12
	1960–94	-1.95 ± 0.15 (–0.53)	-1.04 ± 0.06 (–0.10)	-0.33 ± 0.33	0.53 ± 0.26

Continental trends (and fitted standard errors) for shortwave and net radiation and precipitation forcing data used in the ALL run, and the runoff are shown. Shortwave radiation values in brackets do not include direct and first-indirect aerosol effects (that is, CLIM-only trends). Positive radiation trends indicate a surface energy increase. Units are in $\text{kg m}^{-2} \text{yr}^{-2}$. To directly compare water and energy availability changes for evaporation, radiation terms are converted by considering the amount of energy required to convert 1 kg of liquid water to vapour ($2.5 \times 10^6 \text{ J kg}^{-1}$). Hence, $12.6 \text{ kg m}^{-2} \text{yr}^{-1}$ of water is equivalent to 1 W m^{-2} .

changes in observed runoff. Expanding irrigation reduces continental-scale runoff⁷, particularly over Asia and Europe, which has the largest fractional growth between 1960 and 1990 (ref. 18). Indeed, a majority of global human water consumption is used for irrigated agriculture³. If we assume that most human extraction is from short-residence-time sources, estimates of total human water usage³ would have contributed to a decrease of about $0.2 \text{ kg m}^{-2} \text{yr}^{-2}$ between 1960 and 1995. By combining urban population growth¹⁹ with a global urban map²⁰ and assuming all urban precipitation is lost as runoff, we estimate the global urban runoff contribution to be $\sim +0.03 \text{ kg m}^{-2} \text{yr}^{-2}$. Overall, these additional terms would have contributed to a net reduction in observed runoff trends and therefore cannot explain the difference between observed and climate-driven simulated runoff.

The detection of a direct CO_2 effect on global river runoff is important because, although laboratory experiments have shown that the stomatal openings of many plant species reduce under elevated CO_2 (ref. 5), it was unclear whether this reduction would have any significance for the global water cycle, in which real ecosystems are typically limited by water and nutrient availability. Both atmospheric boundary layer feedbacks²¹ and changes in leaf area index²² have also been proposed as compensating mechanisms that could lead to the negligible impact of CO_2 -induced stomatal closure on large-scale evapotranspiration. As a result, most projections of future water availability have tended to neglect stomatal-closure effects²³. By contrast, our analysis suggests that raised CO_2 levels are already having a direct influence on the water balance at the land surface. As the direct CO_2 effect reduces surface energy loss due to evaporation, it is likely to add to surface warming^{24,25} as well as increasing freshwater availability. The existence of a direct CO_2 signal in river runoff records also opens up the intriguing possibility of using long-term river records to monitor CO_2 effects on the land carbon sink.

METHODS

Monthly forcing data. The 0.5° resolution observational data set from the Climate Research Unit contains monthly temperature (mean and diurnal range), humidity, cloud cover and precipitation (amount and daily frequency)⁹. Climatological wind speed and surface pressure fields are taken from a Hadley Centre global climate model simulation of HadCM3 (ref. 26). The empirical formulations of Albrecht²⁷ are used to derive surface downward shortwave and longwave radiation from the Climate Research Unit data set. These radiation components are compared to the monthly mean radiation data from the Second Global Soil Wetness Project²⁸, which is a combination of reanalysis and observations. The average difference between the two is less than 5 W m^{-2} for both longwave and shortwave radiation, implying that these formulations are sufficient for this purpose.

As we are using cloud-cover observations in the reconstruction of surface downward shortwave radiation, a component of the second indirect aerosol effect (cloud lifetime) is already incorporated. Whether the enhanced cloudiness observed (for example, over North America²⁹) is due to aerosols cannot, however, be ascertained. We use diagnostics from the Met Office HadGEM1

(ref. 13) global climate model to incorporate other changes in surface shortwave radiation due to aerosol. These include the direct (clear-sky scattering and absorption) and indirect effects of the following aerosol species: ammonium sulphate, fossil-fuel black carbon (direct effect only), biomass-burning aerosol and sea salt. We can therefore estimate the long-term trends in direct and first-indirect (cloud albedo) effects and modify the reconstructed clear-sky shortwave radiation and cloud albedo accordingly.

Table 1 shows the regional trends in the data used to force MOSES over the past century and after 1960. (These 'continental' values are only based on catchments with gauged rivers.) Except over North America, the reduction in reconstructed shortwave radiation is mainly due to the direct and first-indirect (cloud albedo) aerosol effects, rather than cloud cover. This is partially offset by the increasing downward longwave radiation (not shown) resulting in a reduction in net radiation at the surface.

Model set-up. All the monthly forcing data are regridded onto the HadCM3 $2.5^\circ \times 3.75^\circ$ grid and disaggregated to hourly data. A half-sine wave is used to calculate the shortwave radiation during the daytime and a sine wave is used to represent the daily temperature cycle. We use a random number generator to specify when a rainfall event occurs. A representative precipitation duration has been set for large-scale and convective precipitation events by analysing HadCM3 runs. (Modelled continental runoff anomalies are found to be insensitive to a range of physically reasonable precipitation event durations.) The MOSES soil water and temperature are 'spun up' by repeatedly forcing the model with monthly data averaged over the period 1901–1905 and setting CO_2 and land cover to 1901 values. MOSES is then run over the 1901–1994 period. The variation in total annual runoff over a number of regions is compared with the observationally based data set¹. We only analyse the period during which the regional data¹ are based on observations from at least 20% of the total river basin area (see Table 1 for details), as below this threshold we find a large drop in the correlation between observed precipitation and runoff (not shown). Our catchment area estimates are from a different (digital) source³⁰, so we rescale our total continental runoff by the ratio of the sum of catchment areas used in each study (corresponding to a 5% change globally). The modelled global mean runoff for the 1901–1994 study period is $3.92 \times 10^4 \text{ km}^3 \text{yr}^{-1}$ as compared to $3.94 \times 10^4 \text{ km}^3 \text{yr}^{-1}$ (ref. 1).

Received 5 July; accepted 7 December 2005.

1. Labat, D., Godderis, Y., Probst, J. L. & Guyot, J. L. Evidence for global runoff increase related to climate warming. *Adv. Water Res.* **27**, 631–642 (2004).
2. Probst, J. L. & Tardy, Y. Long range stream-flow and world continental runoff fluctuations since the beginning of this century. *J. Hydrol.* **94**, 289–311 (1987).
3. Shiklomanov, I. A. Appraisal and assessment of world water resources. *Water Int.* **25**, 11–32 (2000).
4. Stanhill, G. & Cohen, S. Global dimming: a review of the evidence for a widespread and significant reduction in global radiation with discussion of its probable causes and possible agricultural consequences. *Agric. For. Met.* **107**, 255–278 (2001).
5. Field, C., Jackson, R. & Mooney, H. Stomatal responses to increased CO_2 : implications from the plant to the global-scale. *Plant Cell Environ.* **18**, 1214–1255 (1995).
6. Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M. & Taylor, C. M. Explicit representation of sub-grid heterogeneity in a GCM land surface scheme. *J. Hydromet.* **4**, 530–543 (2003).
7. Tett, S. F. B. *et al.* Estimation of natural and anthropogenic contributions to 20th century temperature change. *J. Geophys. Res.* **107**, 4306, doi:10.1029/2000JD000028 (2002).

8. Roderick, G. D. & Farquhar, G. D. The cause of decreased pan evaporation over the last 50 years. *Science* **298**, 1410–1411 (2002).
9. New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **13**, 2217–2238 (2000).
10. Cox, P. M., Huntingford, C. & Harding, R. J. A canopy conductance and photosynthesis model for use in a GCM land surface scheme. *J. Hydrol.* **213**, 79–94 (1998).
11. Harding, R. J., Huntingford, C. & Cox, P. M. Modelling long-term transpiration measurements from grassland in southern England. *Agric. For. Met.* **100**, 309–322 (2000).
12. Pope, V. D., Gallani, M. L., Rowntree, P. R. & Stratton, R. A. The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Clim. Dyn.* **16**, 123–146 (2000).
13. Stott, P. A. *et al.* HadGEM1—Transient simulations with HadGEM1 using historic and SRES scenarios. *J. Clim.* (in the press).
14. Goldewijk, K. K. Estimating global land use change over the past 300 years: the HYDE database. *Glob. Biogeochem. Cycles* **15**, 417–433 (2001).
15. Mitchell, J. F. B. & Karoly, D. J. in *Climate Change 2001. Impacts: The Scientific Basis* (eds Houghton, J. T. *et al.*) 697–738 (Cambridge Univ. Press, Cambridge, 2001).
16. Lambert, F. H., Stott, P. A., Allen, M. R. & Palmer, M. A. Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.* **31**, L10203, doi:10.1029/2005GL023654 (2004).
17. Boucher, O., Myhre, G. & Myhre, A. Direct human influence of irrigation on atmospheric water vapour and climate. *Clim. Dyn.* **22**, 597–603 (2004).
18. FAOSTAT Statistical Databases: Land Use. (<http://faostat.fao.org/>) (Food and Agriculture Organisation of the United Nations, July 2005).
19. FAOSTAT Statistical Databases: Population. (<http://faostat.fao.org/>) (Food and Agriculture Organisation of the United Nations, July 2005).
20. Loveland, T. R. *et al.* Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. *Int. J. Remote Sens.* **21**, 1303–1330 (2000).
21. McNaughton, K. G. & Jarvis, P. G. Effects of spatial scale on stomatal control of transpiration. *Agric. For. Met.* **54**, 279–302 (1991).
22. Betts, R. A., Cox, P. M., Lee, S. E. & Woodward, F. I. Contrasting physiological and structural vegetation feedbacks in climate change simulations. *Nature* **387**, 796–799 (1997).
23. Arnell, N. & Lui, C. in *Climate Change 2001. Impacts: Adaptation and Vulnerability* (eds McCarthy, J. J. *et al.*) 191–234 (Cambridge Univ. Press, Cambridge, 2001).
24. Sellers, P. J. *et al.* Comparison of radiative and physiological effects of doubled atmospheric CO₂ on climate. *Science* **271**, 1402–1406 (1996).
25. Cox, P. M. *et al.* The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).
26. Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16**, 147–168 (2000).
27. Henning, D. in *Atlas of the Surface Heat Balance of the Continents* 6–9 (Gebrüder Borntraeger, Berlin, 1989).
28. Zhao, M. & Dirmeyer, P. Production and analysis of GSWP2 near-surface meteorology data sets. *COLA Technic. Rep.* **159**, 1–36 (2003).
29. Liepert, B. G. Observed reductions of surface solar radiation at sites in the United States and worldwide from 1961 to 1990. *Geophys. Res. Lett.* **29**, 1421, doi:10.1029/2002GL014910 (2002).
30. Fekete, B. M., Vorosmarty, C. J. & Grabs, W. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Glob. Biogeochem. Cycles* **16**, 1042, doi:10.1029/1999GB001254 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Labat (Laboratoire de Mécanisme de Transferts en Géologie, UMR CNRS, Toulouse, France) for providing the observational runoff data. We thank D. Sexton for statistical advice and A. Jones for discussions. N.G., R.A.B., O.B. and P.A.S. were supported by the UK Department for Environment, Food and Rural Affairs, and P.M.C. and C.H. by the UK Natural Environment Research Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.G. (nicola.gedney@metoffice.gov.uk).

Skull of the large non-macrostromatan snake *Yurlunggur* from the Australian Oligo-Miocene

John D. Scanlon^{1,2}

Understanding the origin and early evolution of snakes from lizards depends on accurate morphological knowledge of the skull in basal lineages, but fossil specimens of archaic snakes have been rare, and either fragmentary or difficult to study as a result of compression by enclosing sediments^{1–6}. A number of Cenozoic fossil snakes from Australia have vertebral morphology diagnostic of an extinct group, Madtsoiidae, that was widespread in Gondwana from mid-Cretaceous (Cenomanian) to Eocene times, and also reached Europe in the late Cretaceous period^{3,7–11}. Despite this long history, only about half the skull is known from the best-known species *Wonambi naracoortensis*^{7,11–13}, and the few known cranial elements of other species have added little further evidence for phylogenetic relationships^{10,14–19}. Conflicting hypotheses have been proposed for their relationships and evolutionary significance, either as basal ophidians with many ancestral (varanoid- or mosasaur-like) features, or advanced (macrostromatan) alethinophidians of little relevance to snake origins^{3,4,7,11–15,20}. Here I report two partial skeletons referred to *Yurlunggur*⁸, from the late Oligocene and early Miocene of northern Australia, which together represent almost the complete skull and mandible. The exceptionally preserved skulls provide new evidence linking *Yurlunggur* with *Wonambi* and other madtsoiids, falsifying predictions of the macrostromatan hypothesis, and supporting the exclusion of Madtsoiidae from the clade including all extant snakes.

Freshwater carbonate deposits of the Riversleigh World Heritage Area, northwestern Queensland, preserve rich samples of successive vertebrate local faunas from about 25–12 Myr ago²¹. Riversleigh fossils are rarely articulated but usually undistorted, can be completely freed from matrix with acid, and include the most diverse known assemblages of Madtsoiidae^{8,10,11,16}. This group comprises small to very large snakes (total adult length estimated to range from 0.5 to 6.5 m in Australian forms, and to more than 9 m elsewhere), presumed to be constrictors ecologically analogous to living pythons and boas^{3,7–13,22,23}.

Squamata Oppel, 1811
Ophidia Brongniart, 1800
Madtsoiidae Hoffstetter, 1961
Yurlunggur Scanlon, 1992

Stratigraphic and geographic range. Late Oligocene to Late Pleistocene; northeastern and central Australia (Queensland, Northern Territory, South Australia, New South Wales^{8,16–18,23}).

Remarks. Madtsoiid monophyly is only weakly supported as long as most included taxa are inadequately known^{13,18,19}; their taxonomy has been based on vertebrae, and the single named species of *Yurlunggur* (*Y. camfieldensis*) is known only from postcranial remains^{8,23}. Revised diagnoses of Madtsoiidae and *Yurlunggur* are provided in Supplementary Information; in the Riversleigh faunas, vertebrae attributable to *Yurlunggur* are readily distinguished from

other co-occurring madtsoiids (including species of *Wonambi* and *Nanowana*), pythons, typhlopids and elapids by numerous shape characters as well as much larger adult size^{8,10,11,16,24,25}. Some isolated skull elements can also be referred to this taxon on the basis of their large size^{16–18}. More than one new species may be represented in the material reported here, but pending further study of vertebrae the new skeletons are simply referred to *Yurlunggur* sp. or spp.

Referred material. Hiatus A Site (Late Oligocene²¹): disarticulated remains of two different-sized adults in blocks collected in 2002, the larger including more than 40 cervical, trunk, cloacal and caudal vertebrae, ribs, fragments of the right maxilla (Queensland Museum F45217), dentary (QMF45073) and frontal (QMF45388), and articulated partial braincase (QMF45111, Fig. 1i–m) comprising prootics, exoccipital–opisthotics, basioccipital, partial supraoccipital and part of the parietal. Smaller individual represented by right frontal (QMF45389) and exoccipital–opisthotic (QMF45390) as well as several trunk vertebrae inconsistent (in combination of size and regional features) with the rest of the material. CS Site (Early Miocene²¹): associated and partly articulated skeleton (QMF45391) of a small adult in several contiguous blocks collected in 2000, including most of the skull (premaxilla, nasals, septomaxillae, vomers, maxillae, palatines, pterygoid, ectopterygoids, frontals, prefrontals, postorbitofrontals, jugals, parietal, supraoccipital, parabasisphenoid, quadrates, splenials and dentary; Fig. 1a–h) and more than 60 vertebrae including atlas, axis and several successive cervicals, articulated series of mid-trunk vertebrae with ribs, and additional preloacal, cloacal and caudal vertebrae. Disarticulated vertebrae of several other (larger and smaller) *Yurlunggur* individuals were recovered from the same blocks.

Measurements of overlapping parts of the skulls (frontal, supraoccipital, parietal–prootic contacts) show that QMF45111 has linear dimensions about 1.3-fold those of QMF45391. Maximum vertebral sizes show a similar ratio of width across prezygapophyses (41.0 mm:33.2 mm $\approx$ 1.2), and a simple proportionality method^{9,23} gives total length estimates of 5.7 and 4.7 m respectively. The description below is based principally on the two most complete skeletons, but other referred cranial material from Riversleigh includes the posterior part of the parabasisphenoid (MM Site QMF23041) and articular region of the mandible (BSE Site QMF23060) as well as jaw elements represented in QMF45391 (refs 13, 16–19).

Most significant new morphological information. The premaxilla of QMF45391 lacks teeth or alveoli and has concave lateral facets indicating tight contact with the maxillae, consistent with a mobile articulation but not the loose ligamentous connection found in most modern snakes¹⁵. The maxilla, palatine, pterygoid and dentary resemble those of *Nanowana* spp. more than those of *Wonambi*^{10,11,13,19}; the complete maxilla has 22 alveoli, palatines 10/11, pterygoid 7 (but 9 in QMF51378; WH Site), and dentary

¹Riversleigh Fossil Centre, Outback at Isa, PO Box 1094, Mount Isa, Queensland 4825, Australia. ²South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia.

(incomplete posteriorly) at least 20. The maxillary dentition is nearly isodont, dentary is slightly proterodont, and radial infolding of the tooth bases ('picidentine'^{11,17}) is observable on broken or detached teeth in both new specimens. The nasals, septomaxillae and vomers are similar to those of *Cylindrophis* except that the nasals are narrow and separated from the prefrontals by posterior extensions of the external nares reaching the frontals, as in varanoid and mosasauroid lizards but not *Dinilysia* or any basal modern snakes¹⁵. Posteriorly, the nasals extend slightly between the frontals mesially and overlap them dorsally, probably forming a somewhat mobile prokinetic articulation as in basal alethinophidians, not a strong suture as inferred in *Wonambi*^{11,13,26}. The prefrontal has interlocking but potentially mobile articulations with both the frontal and the distinct dorsal process of the maxilla, possibly permitting rocking motions but little or no sliding at the prefrontal–maxillary contact, and thus resembles anilioids but neither macrostomatans nor *Wonambi*^{11,13,15,26,27}. The orbital lamina of the prefrontal has two deep, ventrally open notches representing the foramina of the single lacrimal duct (between prefrontal and maxilla) and palatine nerve (between prefrontal and palatine); there is no horizontal lamina roofing the lacrimal duct, as present in macrostomatans apart from some booids¹⁵. The frontals

have descending interolfactory processes, their anterior surfaces forming a clasping contact with the nasals, but the undamaged subolfactory laminae have no ascending projections and hence the interolfactory pillars are incomplete; complete (sutured or fused) pillars are synapomorphic for Alethinophidia, with no instances of reversal known^{1,13,15,20}. The parietal is relatively both shorter and narrower than in *Wonambi* and *Dinilysia*, and the transverse anterior margin and anterolateral (postorbital) processes extend further anteriorly to underlap and clasp the frontals, thus forming supra-orbital processes as in anilioids, xenopeltids and *Haasiophis*^{1,5,6,13,15}. The postorbital bar is complete, with two triradiate elements each forming about half of the posterior border of the orbit, unlike all modern snakes¹⁵ but similar to *Dinilysia*⁵. The stepped, oblique overlap of the upper and lower elements is comparable to the plesiomorphic postorbital–jugal contact retained in many lizards²⁸, suggesting that the bones can be identified respectively as postorbito-frontal (fused postfrontal and postorbital, contacting the frontal and prefrontal, and clasping the anterolateral margin of the parietal supraorbital process) and jugal (forming an elongate, narrow contact with a dorsal facet on the maxilla); this raises questions concerning homology of the 'postorbital' in modern snakes, which might

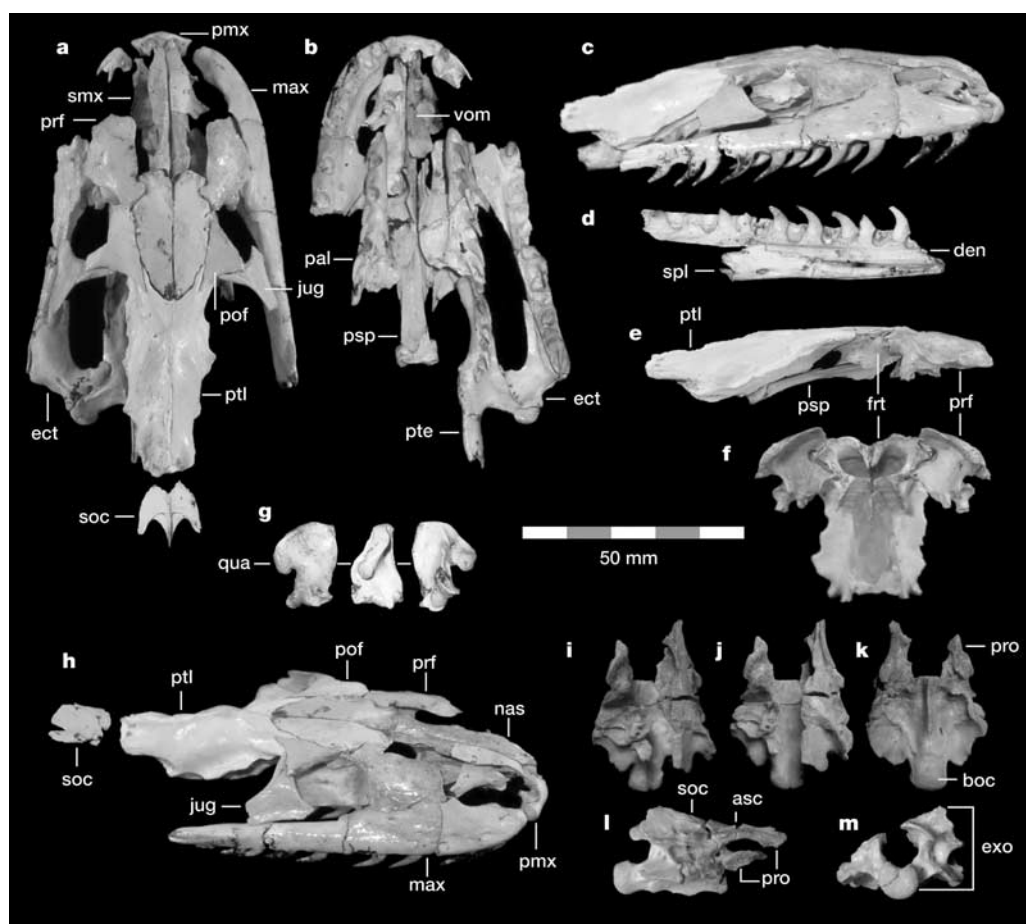


Figure 1 | *Yurlunggur* sp. or spp. (Madtsoiidae), parts of two skulls from Riversleigh, Queensland, Australia. **a–h**, QMF45391 (CS Site, Early Miocene); **i–m**, QMF45111 (Hiatus A Site, Late Oligocene). **a–c**, Dorsal (**a**), palatal (**b**) and right lateral (**c**) views of rearticulated cranial bones; supraoccipital shown in **a** only; parietal and posterior part of right maxilla omitted in **b**. **d**, Medial view of left dentary (omitting anterior portion) and splenial. **e**, **f**, Right lateral (**e**) and anteroventral (**f**) views of parietal, parasphenoid, frontals and prefrontals (same elements partly obscured in **a–c**). **g**, Right quadrate in lateral, posterior and medioventral views. **h**, The same elements as **a**, in right dorsolateral and slightly anterior view.

i–m, Braincase elements in left dorsolateral (**i**), dorsal (**j**), ventral (**k**), right lateral (**l**) and posterior (**m**) views (parts of left prootic and parietal omitted). Abbreviations: asc, furthest extent of anterior semicircular canal (exposed in repaired crack); boc, basioccipital; den, dentary; ect, ectopterygoid; exo, exoccipital–opisthotic; frt, frontal; jug, jugal; max, maxilla; nas, nasal; pal, palatine; pmx, premaxilla; pof, postorbitofrontal; prf, prefrontal; pro, prootic; psp, parasphenoid; pte, pterygoid; ptl, parietal; qua, quadrate; smx, septomaxilla; soc, supraoccipital; spl, splenial; vom, vomer. Scale bar, 50 mm.

represent or incorporate the actual jugal (usually considered absent^{5,10–13,15,28}). The parasphenoid broadly enters the optic fenestra, as in *Dinilysia* and *Wonambi* but no modern snakes except colubroids, which are deeply nested within Macrostromata^{1,13,20}. The strongly arched choanal process of the palatine wraps around the posterior process of the vomer, and abuts the frontal descensus and a sagittal keel (interchoanal process) of the parasphenoid rostrum, consistent with slight kinesis (lateral rotation) of the palatine relative to the braincase, as in anilioids^{26,27}. The ectopterygoid clasps the large lateral process of the pterygoid both dorsally and ventrally, as in most lizards but unlike any modern snake (this region is poorly known in other fossil snakes)^{11,15,16}. The quadrate is similar in shape to those of *Dinilysia* and some basal alethinophidians (particularly *Cylindrophis* and *Xenopeltis*): nearly vertical and relatively short dorsoventrally, with a prominent and downcurved suprastapedial process^{1,5,15,26}. There is a distinct stapedial facet on the posteroventral surface of the process, but as in *Dinilysia* and *Xenopeltis* the extracolumella of the stapes (stylohyal) is not fused either to the process as in anilioids or to the shaft of the quadrate as in other macrostomatans.

The prootic (QMF45111) has a large, undivided trigeminal notch, without damage indicating that a laterosphenoid bridge was ever present on either side (as in *Wonambi* and *Dinilysia*, most simply interpreted as absence of the alethinophidian synapomorphy^{1,5,11–13,15,20}). As in the latter fossils the anterodorsal (alar) process is elongate and shallow, but here it is confirmed to extend far beyond the anterior semicircular canal, unlike in all modern snakes, that have the crista alaris greatly reduced and the labyrinth extending close to the anterior margin of the prootic^{13,15,29}. The left and right exoccipitals meet broadly on the dorsal surface of the occipital condyle (as in *Wonambi*, *Dinilysia*, *Cylindrophis* and *Anomochilus*; separated in most other squamates^{1,13,15,27}), and narrowly above the foramen magnum. In QMF45111, the exoccipital–opisthotics are partly fused with the basioccipital and supraoccipital, whereas the supraoccipital (QMF45391) and exoccipital–opisthotic (QMF45390) are unfused in both smaller specimens, suggesting that fusion occurred in late ontogeny (rare in modern squamates).

The dentary (most complete in QMF45391) is relatively narrow but widest posteriorly, with two elongate mental foramina on the

lateral face below alveoli 3–4 and 7–8. The large splenial has a long interlocking contact with the medioventral crest of the dentary, a concave posterior articular surface for the angular (intramandibular joint) and a dorsoposterior facet for an anterior process of the coronoid.

Yurlunggur, hitherto known only from few isolated elements^{10,16–19}, is now more completely known than *Wonambi naracoortensis*^{7,11–13}. Almost every cranial element is now represented in one or both of these taxa; only the supratemporal, stapes, angular, coronoid and epipterygoid (if present¹³) have not yet been identified. Of 212 osteological characters used in a recent phylogenetic analysis¹⁵, *Yurlunggur* can now be scored for 182, *Wonambi* (including *W. barriei*^{11,16}) for 143, and either or both for 188.

Previous analyses inferring Madtsoiidae to be archaic, non-alethinophidian snakes^{11,15,18,20} are further supported by characters newly discovered in *Yurlunggur*, many of which have never previously been observed in fossil snakes. The competing hypothesis, that *Wonambi* and related taxa belong to the extant clade Macrostromata^{4,7,12}, would predict the presence of apomorphies including complete interolfactory pillars, laterosphenoid dividing trigeminal notch, crista alaris of prootic reduced relative to semicircular canals, extracolumella of stapes fusing to the quadrate, prefrontal roofing lacrimal duct, parasphenoid excluded from optic fenestra, and reduced ectopterygoid process overlapped only dorsally or laterally by ectopterygoid^{5,6,12,13,15,20}. *Yurlunggur* possesses ancestral or intermediate states for each of these characters, previously unobserved in madtsoiids or disputed in *Wonambi*. These two Australian snakes differ in many details but are related (forming a clade with bootstrap support of 82%; Fig. 2) and apparently more distant from modern snakes than is *Dinilysia* (72%). There is strong support for monophyly of the nested clades Serpentes, Alethinophidia and Macrostromata (98%, 62% and 99%, respectively) as excluding both of the Cenozoic madtsoiids as well as the three best-known Cretaceous ophidians. These long-jawed or 'macro-stomate' fossils are thus excluded from Macrostromata (which is diagnosed by numerous characters, not all related to the jaw apparatus¹⁵), and collectively represent the ancestral morphology of the diverse modern snake radiation.

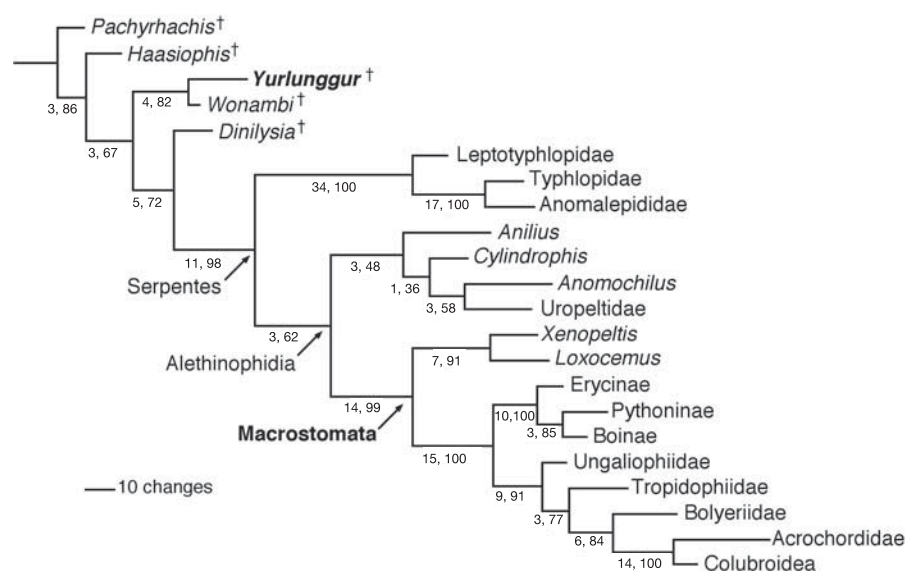


Figure 2 | Phylogenetic relationships of extant and extinct snakes (Ophidia). The cladogram is based on a recent phenotypic data matrix¹⁵, modified as explained in Supplementary Information. The single most parsimonious cladogram was found by using PAUP³⁰ (with multistate characters treated as numerically or topologically ordered where reasonable, and a composite 'varanoid' outgroup¹⁵), and is shown with branch lengths

proportional to the minimum number of changes; cladistically uninformative characters are excluded, so only homoplastic changes are shown on terminal branches. Measures of clade support (Bremer decay index, bootstrap percentage) are shown on branches. Extinct taxa are marked (†), and clade names Serpentes (crown group of 'modern snakes'), Alethinophidia and Macrostromata are placed at their basal nodes.

Received 20 May; accepted 17 August 2005.

- Estes, R., Frazzetta, T. H. & Williams, E. E. Studies on the fossil snake *Dinilysia patagonica* Woodward: Part 1. Cranial morphology. *Bull. Mus. Comp. Zool. Harvard* **140**, 25–74 (1970).
- Lee, M. S. Y. & Caldwell, M. W. Anatomy and relationships of *Pachyrhachis problematicus*, a primitive snake with hindlimbs. *Phil. Trans. R. Soc. Lond. B* **353**, 1521–1552 (1998).
- Rage, J.-C. & Werner, C. Mid-Cretaceous (Cenomanian) snakes from Wadi Abu Hashim, Sudan: The earliest snake assemblage. *Palaeontol. Afr.* **35**, 85–110 (1999).
- Coates, M. & Ruta, M. Nice snake, shame about the legs. *Trends Ecol. Evol.* **15**, 503–507 (2000).
- Caldwell, M. W. & Albino, A. Exceptionally preserved skeletons of the Cretaceous snake *Dinilysia patagonica* Woodward, 1901. *J. Vertebr. Paleontol.* **22**, 861–866 (2002).
- Rieppel, O., Zaher, H., Tchernov, E. & Polcyn, M. J. The anatomy and relationships of *Haasiophis terrasanctus*, a fossil snake with well-developed hind limbs from the mid-Cretaceous of the Middle East. *J. Paleontol.* **77**, 536–558 (2003).
- Barrie, D. J. Skull elements and associated remains of the Pleistocene boid snake *Wonambi naracoortensis*. *Mem. Queensl. Mus.* **28**, 139–151 (1990).
- Scanlon, J. D. A new large madtsoiid snake from the Miocene of the Northern Territory. *Beagle Rec. N. Territory Mus. Arts Sci.* **9**, 49–60 (1992).
- Scanlon, J. D. Madtsoiid snakes from the Eocene Tingamarra Fauna of eastern Queensland. *Kaupia: Darmst. Beitr. Naturgeschichte* **3**, 3–8 (1993).
- Scanlon, J. D. *Nanowana* gen. nov., small madtsoiid snakes from the Miocene of Riversleigh: sympatric species with divergently specialised dentition. *Mem. Queensl. Mus.* **41**, 393–412 (1997).
- Scanlon, J. D. & Lee, M. S. Y. The Pleistocene serpent *Wonambi* and the early evolution of snakes. *Nature* **403**, 416–420 (2000).
- Rieppel, O., Kluge, A. G. & Zaher, H. Testing the phylogenetic relationships of the Pleistocene snake *Wonambi naracoortensis* Smith. *J. Vertebr. Paleontol.* **22**, 812–829 (2002).
- Scanlon, J. D. Cranial morphology of the Plio-Pleistocene giant madtsoiid snake *Wonambi naracoortensis*. *Acta Palaeontol. Pol.* **50**, 139–180 (2005).
- Rage, J.-C. Fossil snakes from the Paleocene of São José de Itaboraí, Brazil. Part I. Madtsoiidae, Aniliidae. *Palaeovertebrata* **27**, 109–144 (1998).
- Lee, M. S. Y. & Scanlon, J. D. Snake phylogeny based on osteology, soft anatomy and behaviour. *Biol. Rev.* **77**, 333–402 (2002).
- Scanlon, J. D. *Studies in the Palaeontology and Systematics of Australian Snakes* Thesis, Univ. New South Wales (1996).
- Scanlon, J. D. & Lee, M. S. Y. On varanoid-like dentition in primitive snakes (Madtsoiidae). *J. Herpetol.* **36**, 100–106 (2002).
- Scanlon, J. D. The basicranial morphology of madtsoiid snakes (Squamata, Ophidia) and the earliest Alethinophidia (Serpentes). *J. Vertebr. Paleontol.* **23**, 971–976 (2003).
- Scanlon, J. D. Australia's oldest known snakes: *Patagoniophis*, *Alamitophis*, and cf. *Madtsoia* (Squamata: Madtsoiidae) from the Eocene of Queensland. *Mem. Queensl. Mus.* **51**, 215–235 (2005).
- McDowell, S. B. in *Snakes: Ecology and Evolutionary Biology* (eds Seigel, R. A., Collins, J. T. C. & Novak, S. S.) 1–50 (Macmillan, New York, 1987).
- Archer, M., Hand, S. J., Godthelp, H. & Creaser, P. Correlation of the Cainozoic sediments of the Riversleigh World Heritage fossil property, Queensland, Australia. *Mém. Trav. École Prat. Hautes Études Inst. Montpellier* **21**, 131–152 (1997).
- Simpson, G. G. A new fossil snake from the *Notostylops* beds of Patagonia. *Bull. Am. Mus. Nat. Hist.* **67**, 1–22 (1933).
- Scanlon, J. D. First known axis vertebra of a madtsoiid snake (*Yurlunggur camfieldensis*) and remarks on the neck of snakes. *Beagle Rec. N. Territory Mus. Arts Sci.* **20**, 207–215 (2004).
- Scanlon, J. D. *Montypythonoides*: the Miocene snake *Morelia riversleighensis* (Smith and Plane, 1985) and the geographical origin of pythons. *Mem. Assoc. Australas. Palaeontol.* **25**, 1–35 (2001).
- Scanlon, J. D., Lee, M. S. Y. & Archer, M. Mid-Tertiary elapid snakes (Squamata, Colubroidea) from Riversleigh, northern Australia: early steps in a continent-wide adaptive radiation. *Geobios* **36**, 573–601 (2003).
- Cundall, D. Feeding behaviour in *Cylindrophis* and its bearing on the evolution of alethinophidian snakes. *J. Zool. (Lond.)* **237**, 353–376 (1995).
- Scanlon, J. D. & Rossman, D. S. Cephalic anatomy of the rare Indonesian snake *Anomochilus weberi*. *Zool. J. Linn. Soc.* **109**, 235–273 (1993).
- Estes, R., de Queiroz, K. & Gauthier, J. A. in *Phylogenetic Relationships of the Lizard Families* (eds Estes, R. & Pregill, G.) 119–281 (Stanford Univ. Press, Stanford, California, 1988).
- Rieppel, O. & Zaher, H. The braincases of mosasaurs and *Varanus*, and the relationships of snakes. *Zool. J. Linn. Soc.* **129**, 489–514 (2000).
- Swofford, D. L. *PAUP* Version 4—Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, Massachusetts, 1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank M. Archer, H. Godthelp, A. Gillespie and K. Black (University of New South Wales (UNSW), Sydney) and B. Chambers (Riversleigh Fossil Centre, Mount Isa) for access to and initial preparation of Riversleigh fossils collected by the UNSW team; M. Lee, M. Hutchinson, T. Flannery and J. McNamara (South Australian Museum) for access to facilities and materials; and J.-C. Rage, Z. Szyndlar, M. Lee, M. Kearney, C. Davidson and E. Chitra for discussions and comments. Financial support was provided by the Australian Research Council and the Australian Geographic Society.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to J.D.S. (riversleigh@outbacktisa.com.au).

Positive and negative effects of widespread badger culling on tuberculosis in cattle

Christl A. Donnelly^{1,2}, Rosie Woodroffe^{2,3}, D. R. Cox^{2,4}, F. John Bourne², C. L. Cheeseman⁵, Richard S. Clifton-Hadley⁶, Gao Wei¹, George Gettinby^{2,7}, Peter Gilks¹, Helen Jenkins¹, W. Thomas Johnston¹, Andrea M. Le Fevre^{1†}, John P. McInerney^{2,8} & W. Ivan Morrison^{2,9}

Human and livestock diseases can be difficult to control where infection persists in wildlife populations. For three decades, European badgers (*Meles meles*) have been culled by the British government in a series of attempts to limit the spread of *Mycobacterium bovis*, the causative agent of bovine tuberculosis (TB), to cattle¹. Despite these efforts, the incidence of TB in cattle has risen consistently, re-emerging as a primary concern for Britain's cattle industry. Recently, badger culling has attracted controversy because experimental studies have reached contrasting conclusions (albeit using different protocols), with culled areas showing either markedly reduced^{2,3} or increased^{4,5} incidence of TB in cattle. This has confused attempts to develop a science-based management policy. Here we use data from a large-scale, randomized field experiment to help resolve these apparent differences. We show that, as carried out in this experiment, culling reduces cattle TB incidence in the areas that are culled, but increases incidence in adjoining areas. These findings are biologically consistent with previous studies^{2–5} but will present challenges for policy development.

Bovine tuberculosis is a zoonotic disease with serious consequences for Britain's cattle industry¹. The causative agent, *Mycobacterium bovis*, has a broad host range, which in Britain includes the European badger, a widespread but protected wildlife species¹. Regular testing of cattle herds and slaughterhouse surveillance form the main components of national TB surveillance and control; animals that test positive for TB are slaughtered and affected herds are placed under temporary movement restrictions. Similar measures have eradicated TB in cattle across much of the developed world¹. Patterns of infection in cattle and badgers are closely associated⁶ and, in Britain, cattle-based TB controls have been supplemented with various forms of badger culling^{1,7,8}. Despite these efforts, over the past two decades substantial increases have been recorded in both the incidence and geographic extent of cattle infection, prompting repeated reviews of control options^{1,9,10}.

The development of improved TB control strategies has been hindered by contrasting findings regarding the effectiveness of badger culling as a management tool. Two studies in the Republic of Ireland^{2,3} and one in Britain¹¹ have associated substantial reductions in the incidence of TB in cattle with the near-eradication of badgers from large areas of land (a total of six areas ranging from 104 to 528 km²). In contrast, a field experiment conducted in Britain (the randomized badger culling trial or RBCT) indicated higher

cattle TB incidence in nine study areas (each approximately 100 km²) where badgers were culled locally (average area 5.3 km² per cull) in response to specific cattle TB outbreaks ('reactive' culling) than in nearby areas randomly allocated to a no-culling treatment^{4,5}. The Independent Scientific Group on Cattle TB (ISG, with members F.J.B., C.A.D., D.R.C., G.G., J.P.M., W.I.M. and R.W.) designed and oversees the RBCT. The ISG has tentatively suggested⁴ that this apparently detrimental effect of localized badger culling might be generated by the disruption of badgers' territorial organization at artificially reduced population densities^{12–15}, potentially influencing contact rates with cattle. However, neither this explanation nor the effect itself has been universally accepted^{9,16}.

The RBCT included two culling treatments. Localized 'reactive' culling sought to reduce TB risks to cattle while minimizing the number of badgers culled^{1,17}. 'Proactive' culling^{1,17} involved widespread and repeated culling of badgers across entire trial areas and was included in order to estimate the maximum reduction in cattle TB incidence achievable by culling badgers (subject to animal welfare considerations, non-compliance by land occupiers, and other practical constraints likely to influence the long-term sustainability of a potential culling policy)^{17,18}. The rates of cattle TB incidents ('herd breakdowns') in culled areas were then compared with rates in uncultured 'survey-only' areas^{1,4,5}.

Our analyses reveal that, thus far, the incidence of cattle herd breakdowns has been 19% lower in proactive trial areas than in survey-only areas ($P = 0.005$; 95% confidence interval (CI) 6.2–30% reduction; Fig. 1a). The CI is conservatively inflated (like all CIs reported here) to account for any overdispersion). This result was consistent across ten proactive culling areas, each paired with a survey-only area (the test for overdispersion was not significant; $P = 0.58$). The estimated effect was stronger when measured after the first follow-up cull rather than the initial cull (23% reduction, $P = 0.008$; 95% CI: 6.5–36% reduction). Our analyses revealed no significant change in the effect of culling on breakdown incidence over time (see Supplementary Information). The data suggest that the beneficial effects of culling might increase as one moves deeper inside trial area boundaries (see Supplementary Information), but this trend is not consistent. We will continue to examine this possibility as additional data become available.

We also compared the incidence of cattle herd breakdowns in areas up to 2 km outside proactive and survey-only trial areas (Fig. 2). Parallel ecological studies have revealed reduced population densities

¹Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St. Mary's Campus, Norfolk Place, London W2 1PG, UK. ²Independent Scientific Group on Cattle TB, c/o Department for Environment, Food & Rural Affairs, 1A Page Street, London SW1P 4PQ, UK. ³Department of Wildlife, Fish & Conservation Biology, University of California, One Shields Avenue, Davis, California 95616, USA. ⁴Nuffield College, New Road, Oxford OX1 1NF, UK. ⁵Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK. ⁶Veterinary Laboratories Agency, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, UK. ⁷Department of Statistics and Modelling Science, University of Strathclyde, Richmond Street, Glasgow G1 1XH, UK. ⁸Centre for Rural Research, University of Exeter, Lafrondda House, St German's Road, Exeter EX4 6TL, UK. ⁹Centre for Tropical Veterinary Medicine, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK. [†]Present address: Infectious Disease Epidemiology Unit, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

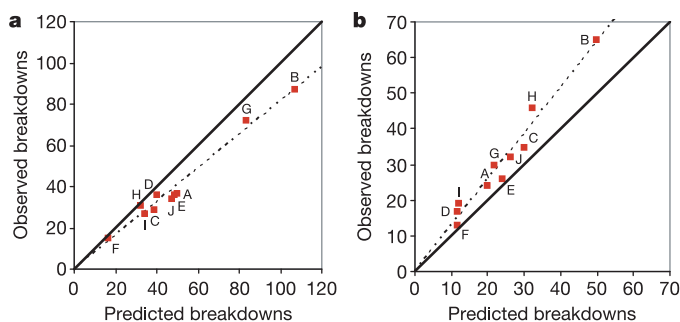


Figure 1 | Effects of badger culling on cattle TB incidence. **a, b,** Observed numbers of confirmed breakdowns (after initial culls), plotted against the numbers of breakdowns predicted with no culling, either inside (**a**) or ≤ 2 km outside (**b**) proactive culling areas A–J. Predictions for each proactive trial area were calculated on the basis of their specific features, replacing the parameter associated with proactive culling with that for survey-only (see Supplementary Information). If incidence rates for proactive areas match those for survey-only areas (after appropriate adjustments), points should fall on the solid lines. Dashed lines indicate the differences in incidence actually observed inside (**a**, total observed, 404; total predicted, 496.7) and outside (**b**, total observed, 307; total predicted, 238.4) trial areas.

and expanded ranging behaviour in badger populations studied up to 2 km outside proactive areas, as well as in reactive areas¹⁹. Therefore, if as hypothesized⁴, expanded badger movement patterns caused the increase in cattle TB incidence recorded in reactive areas, we predicted that a similar pattern would be observed on farms neighbouring proactive trial areas. Analyses revealed a 29% increase in cattle TB incidence ($P = 0.015$; 95% CI: 5.0–58% increase; Fig. 1b) on land neighbouring proactive areas, relative to land neighbouring survey-only areas. Again, the effect was consistent across proactive/survey-only pairs (the test for overdispersion was not significant, $P = 0.38$), but the estimated increase was smaller when measured from the first follow-up cull rather than the initial cull (22% increase, $P = 0.15$; 95% CI: 6.9% reduction to 59% increase). Again, analyses

revealed no significant change in the effect of culling on breakdown incidence over time (see Supplementary Information).

Our results help to resolve apparently contrasting findings from previous studies of badger culling. The 19% reduction in cattle TB incidence observed inside proactive culling areas can be compared with reductions detected, on similar timescales, by two previous studies of widespread culling in Ireland (26% reduction³ and 58% reduction (95% CI: 41–70%)²; see Supplementary Information for details; the different design of a third study¹¹ (in Britain) precluded direct comparison). One explanation for the greater beneficial effect of culling in the Irish studies is that greater reductions in badger density may have been achieved, both because land occupier compliance was higher² and because the culling method (snaring^{2,3}) was probably more efficient than that used in the RBCT (although arguably less publicly acceptable, being perceived as less humane). None of the three previous studies has investigated effects of culling on neighbouring areas. However, in two of these studies^{2,11} we would expect such effects to be weak because culled areas were isolated from neighbouring cattle and badger populations by coastline, rivers or motorways.

Our results are also consistent with findings from the reactive culling treatment of the RBCT. The 29% increase in TB risk that we recorded among cattle herds living close to (but outside) proactive trial areas is similar to the 25% (95% CI: 2.6–52%) increase recorded in reactive areas⁵. This similarity would be expected if proximity to badger culling increases TB risks for cattle (whatever the mechanism) because, in reactive culling areas, small-scale culls were scattered across the landscape, placing a high proportion of herds close to one or more culled areas. The mechanism by which proximity to badger culling may influence cattle TB risks is uncertain. However, the ecological patterns observed are consistent with the hypothesis that culling affects the probability of contact between infected badgers and cattle herds, because badgers range more widely when their densities are artificially reduced^{12–15,19}. Such disruption would be expected to occur throughout the period of a culling strategy, for in the absence of substantial geographic barriers, badgers continue to immigrate into culling areas³. However, contacts with cattle might be particularly frequent at the onset of culling, when territorial

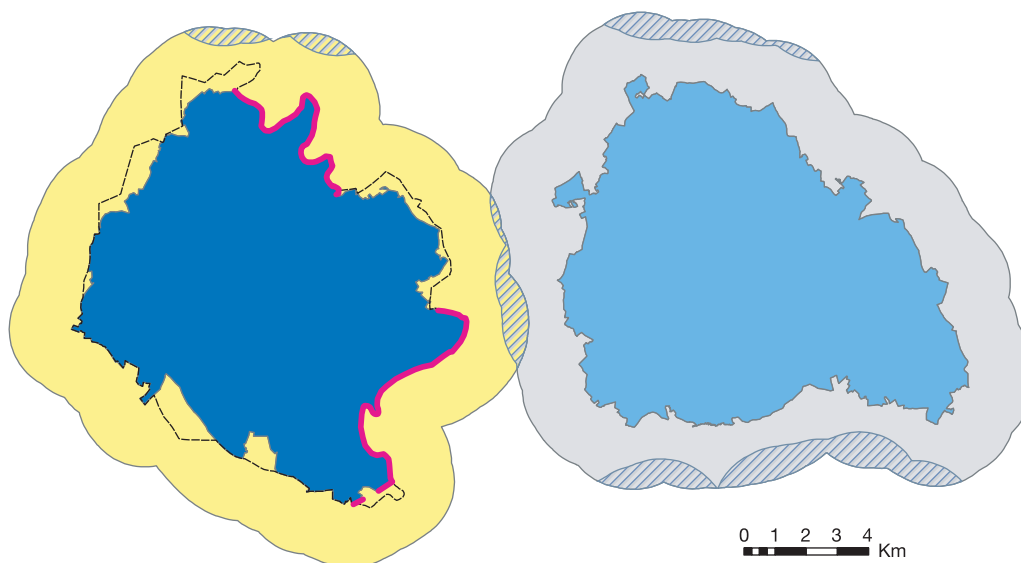


Figure 2 | Simplified map of proactive and survey-only areas of triplet A. Shown are (1) trial areas: proactive trial area (95.7 km², dark blue), survey-only trial area (99.4 km², light blue) and (2) neighbouring areas: proactive neighbouring area (98.9 km², yellow), survey-only neighbouring area (88.0 km², grey). The boundary of the badger-culling area (black dashed line) and possible geographical barriers to badger

movement (red lines) are also shown. Portions of the neighbouring areas overlapping with the neighbouring areas of other nearby trial areas (not shown) account for 4.6 km² of the proactive neighbouring area and 14.9 km² of the survey-only neighbouring area, and are denoted by diagonal hatching.

organization is disrupted but badger population densities may still be comparatively high.

Our finding that widespread culling of badgers has simultaneous positive and negative effects on the incidence of TB in cattle has important implications for the development of sustainable control policies. We would expect the overall reduction in cattle TB to be greatest for very large culling areas (with consequently lower perimeter:area ratios), although in absolute terms the costs, as well as the benefits, will be greatest for large areas. Detailed consideration is needed to determine whether culling on any particular scale would be economically and environmentally sustainable.

METHODS

Trial design. Thirty trial areas were recruited sequentially as ten matched 'triplets' denoted A–J. Trial areas were selected on the basis of high cattle TB incidence. Attempts were made to place trial area boundaries along geographic features that might impede badger movement (for example, coastline, rivers or major roads; Fig. 2), but in most cases this was impossible and trial area boundaries mainly followed property boundaries. Neighbouring trial areas were separated by buffer zones at least 3-km wide. All trial areas were surveyed for badger activity and then randomly allocated to treatments (except in triplet I, for which security concerns directed a specific allocation) such that each treatment—proactive culling, reactive culling, or no culling ('survey only')—was repeated ten times, once within each triplet.

Immediately after treatment allocation, initial proactive culls were conducted on all land for which consent was given¹⁰ (see Supplementary Information). In the absence of geographical barriers, the boundaries of the area to be culled were delineated (beyond trial area boundaries as necessary; Fig. 2) using field survey data to ensure that culling targeted all badgers likely to use farms inside the trial areas. The method used²⁰ provides a good estimate of the disposition of badger home ranges when based on good survey data²¹.

Badgers were captured in cage traps placed primarily at setts; details of culling protocols have been published previously¹⁴. No trapping occurred between 1 February and 30 April each year, to avoid killing mothers with dependent cubs below ground²². Few badgers sustained trap-related injuries²³, and badger killing (by gunshot) was deemed 'humane' by independent audit²⁴. Initial culls for each proactive trial area were completed between December 1998 and December 2002. 'Follow-up' culls were repeated approximately annually (dates in Supplementary Information). Field surveys showed that badger activity in trial areas changed according to the treatment applied^{19,25}, providing no evidence that treatment comparisons were substantially compromised by illegal culling in survey-only areas.

Once the initial proactive cull was complete, data were collected on cattle TB incidence in and around trial areas, using established veterinary surveillance (see Supplementary Information). Data were available until 4 September 2005, and consisted of accrued totals from 46.6 'triplet years' since initial proactive culls and 34.1 'triplet years' since first follow-up culls (where a 'triplet year' is one triplet observed for 12 consecutive months).

Statistical analysis. We used log-linear Poisson regression to compare the numbers of confirmed breakdowns recorded in trial areas subjected to the proactive and survey-only treatments. As in previous analyses^{4,5}, the regression models adjusted for triplet, the log of the number of baseline herds at risk, and the log of the number of confirmed breakdowns recorded in a three year period before RBCT culling (other time periods investigated are presented in the Supplementary Information). Primary analyses included all breakdowns since completion of the initial proactive cull, but we also performed secondary analyses investigating incidence recorded after completion of the first follow-up cull. We performed these secondary analyses for two reasons. First, some breakdowns early in the trial could represent infections acquired (but not detected) before culling began. This could make treatments appear more similar, leading to underestimation of effects; excluding data before the first follow-up cull avoided this potential problem. Second, badger capture rates (see Supplementary Information) suggested that a higher level of badger removal was achieved after the first follow-up cull. Excluding earlier data therefore indicated the maximum potential benefits of culling inside trial areas, but also risked underestimating any detrimental effects in neighbouring areas, which were expected to be greatest immediately after initial culls.

Cattle herd locations from the animal health information system VETNET were used to identify herds inside trial areas. Parallel analyses were performed using the RBCT database to identify herd locations; the results obtained were similar but less consistent (see Supplementary Information). Herds up to 2 km outside trial area boundaries could be identified comprehensively only using

VETNET, because the RBCT database did not include all farms on neighbouring land. Any herds within 2 km of more than one trial area boundary (whether proactive, reactive or survey-only) were omitted from these analyses (Fig. 2). Analyses based on VETNET herd locations include 2,810 herds inside trial areas and 2,163 herds in neighbouring areas.

We sought evidence that the effects of proactive culling were systematically influenced by particular circumstances of local cattle and badger populations, land occupier consent, time since enrolment in the RBCT, and the geographic isolation of trial areas, by investigating interactions of these measures with treatment effects. No such differences were detected in the analyses of confirmed breakdowns (see Supplementary Information).

Models were also fitted replacing the log of the number of baseline herds at risk with either the log of the estimated number of cattle at baseline, the log of the number of tests conducted, or the log of the estimated number of cattle tested, and varying the time period over which past incidence was calculated. The results obtained were similar in each case (see Supplementary Information).

Received 24 October; accepted 28 November 2005.

Published online 14 December 2005.

- Krebs, J. R. *et al.* *Bovine Tuberculosis in Cattle and Badgers* Report PB3423 (HMSO, London, 1997).
- Griffin, J. M. *et al.* The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. *Prev. Vet. Med.* **67**, 237–266 (2005).
- Eves, J. A. Impact of badger removal on bovine tuberculosis in east County Offaly. *Ir. Vet. J.* **52**, 199–203 (1999).
- Donnelly, C. A. *et al.* Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature* **426**, 834–837 (2003).
- Le Fevre, A. M. *et al.* The Impact of Localised Reactive Badger Culling Versus No Culling on TB Incidence in British Cattle: a Randomised Trial. (<http://www.defra.gov.uk/animalh/tb/isg/pdf/lefevre1005.pdf>) (2005).
- Woodroffe, R. *et al.* Spatial association of *Mycobacterium bovis* infection in cattle and badgers. *J. Appl. Ecol.* **42**, 852–862 (2005).
- Dunnet, G. M., Jones, D. M. & McInerney, J. P. *Badgers and Bovine Tuberculosis: Review of Policy* (HMSO, London, 1986).
- Zuckerman, S. *Badgers, Cattle and Tuberculosis* (HMSO, London, 1980).
- Godfray, H. C. J. *et al.* *Independent Scientific Review of the Randomised Badger Culling Trial and Associated Epidemiological Research* (http://www.defra.gov.uk/science/documents/publications/2004/GodfreyReport_BovineTBEpidemiology.pdf) (HMSO, London, 2004).
- Bourne, J. *et al.* *An Epidemiological Investigation into Bovine Tuberculosis—Towards a Science-Based Control Strategy*. Fourth Report of the Independent Scientific Group on Cattle TB (<http://www.defra.gov.uk/animalh/tb/isg/4th-isgreport.pdf>) (Department for Environment, Food & Rural Affairs, London, 2005).
- Clifton-Hadley, R. S., Wilesmith, J. W., Richards, M. S., Upton, P. & Johnston, S. The occurrence of *Mycobacterium bovis* infection in and around an area subject to extensive badger (*Meles meles*) control. *Epidemiol. Infect.* **114**, 179–193 (1995).
- Roper, T. J. & Lups, P. Disruption of territorial behaviour in badgers *Meles meles*. *Z. Säugetierkd.* **58**, 252–255 (1993).
- Cheeseman, C. L., Mallinson, P. J., Ryan, J. & Wilesmith, J. W. in *The Badger* (ed. Hayden, T. J.) 78–93 (Royal Irish Academy, Dublin, 1993).
- O'Corry-Crowe, G., Hammond, R., Eves, J. & Hayden, T. J. The effect of reduction in badger density on the spatial organisation and activity of badgers *Meles meles* L. in relation to farms in central Ireland. *Biol. Env. Proc. R. Ir. Acad.* **96B**, 147–158 (1996).
- Tuytens, F. A. M. *et al.* Spatial perturbation caused by a badger (*Meles meles*) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (*Mycobacterium bovis*). *J. Anim. Ecol.* **69**, 815–828 (2000).
- Woolhouse, M. E. J. Dangers of moving cows. *Nature* **435**, 431–432 (2005).
- Bourne, J. *et al.* *Towards a Sustainable Policy to Control TB in Cattle—A Scientific Initiative*. First Report of the Independent Scientific Group on Cattle TB (Ministry of Agriculture, Fisheries & Food, London, 1998).
- White, P. C. L. & Whiting, S. J. Public attitudes towards badger culling to control bovine tuberculosis in cattle. *Vet. Rec.* **147**, 179–184 (2000).
- Woodroffe, R. *et al.* Effects of culling on badger (*Meles meles*) spatial organization: implications for the control of bovine tuberculosis. *J. Appl. Ecol.* doi:10.1111/j.1365-2664.2005.01144.x (in the press).
- Woodroffe, R., Frost, S. D. W. & Clifton-Hadley, R. S. Attempts to control tuberculosis in cattle by culling infected badgers: constraints imposed by live test sensitivity. *J. Appl. Ecol.* **36**, 494–501 (1999).
- Cresswell Associates. *TB in Cattle: Report of the Independent Audit of Surveying and Badger Social Group Territory Delineation Procedures Used in The Randomised Badger Culling Trial and MAFF's Response*. (<http://www.defra.gov.uk/animalh/tb/publications/pdf/cresswell.pdf>) (2001).
- Woodroffe, R. *et al.* Welfare of badgers (*Meles meles*) subjected to culling: development and evaluation of a closed season. *Anim. Welf.* **14**, 19–25 (2005).

23. Woodroffe, R. *et al.* Welfare of badgers (*Meles meles*) subjected to culling: patterns of trap-related injury. *Anim. Welf.* **14**, 11–17 (2005).
24. Kirkwood, J. K. *Humaneness of MAFF's Badger Despatch Procedures*. (<http://www.defra.gov.uk/animalh/tb/publications/auditor/report.htm>) (2000).
25. Le Fevre, A. M. *et al.* *Changes in Badger Setts Over the First Three Years of the Randomised Badger Culling Trial*. (<http://www.svepm.org.uk/posters/2003/LeFevre.pdf>) (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was funded and implemented by the Department of Environment, Food and Rural Affairs (Defra). We gratefully acknowledge the contribution made by the staff of Defra and its associated

agencies. We also wish to thank the many farmers and land occupiers in the trial areas who allowed the experimental treatments to operate on their land. Finally, we thank D. Mollison for comments.

Author Contributions F.J.B., C.A.D., D.R.C., G.G., J.P.M., W.I.M. and R.W. constitute the Independent Scientific Group on Cattle TB, which designed and oversees the RBCT. Data analyses were designed and performed by C.A.D., D.R.C., G.W., P.G., H.J., W.T.J. and A.M.L.F. The manuscript was drafted by R.W. and C.A.D. with input from all co-authors.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.A.D. (c.donnelly@imperial.ac.uk).

Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea

Niels-Ulrik Frigaard¹†, Asuncion Martinez¹, Tracy J. Mincer¹ & Edward F. DeLong¹

Planktonic Bacteria, Archaea and Eukarya reside and compete in the ocean's photic zone under the pervasive influence of light. Bacteria in this environment were recently shown to contain photoproteins called proteorhodopsins, thought to contribute to cellular energy metabolism by catalysing light-driven proton translocation across the cell membrane^{1–7}. So far, proteorhodopsin genes have been well documented only in proteobacteria and a few other bacterial groups. Here we report the presence and distribution of proteorhodopsin genes in Archaea affiliated with the order Thermoplasmatales, in the ocean's upper water column. The genomic context and phylogenetic relationships of the archaeal and proteobacterial proteorhodopsins indicate its probable lateral transfer between planktonic Bacteria and Archaea. About 10% of the euryarchaeotes in the photic zone contained the proteorhodopsin gene adjacent to their small-subunit ribosomal RNA. The archaeal proteorhodopsins were also found in other genomic regions, in the same or in different microbial lineages. Although euryarchaeotes were distributed throughout the water column, their proteorhodopsins were found only in the photic zone. The cosmopolitan phylogenetic distribution of proteorhodopsins reflects their significant light-dependent fitness contributions, which drive the photoprotein's lateral acquisition and retention, but constrain its dispersal to the photic zone.

Several different groups of Archaea are commonly found in marine plankton^{8,9}. Pelagic crenarchaeotes comprise a large fraction of the prokaryotic plankton in coastal waters and the subphotic zone^{8–14}, and seem to be major participants in chemoautotrophic ammonia oxidation in the ocean^{15–18}. Pelagic euryarchaeotes comprise at least three separate phylogenetic groups based on small-subunit (SSU) rRNA sequence analyses. The most abundant clade, so-called planktonic marine 'group II euryarchaeotes', is peripherally related to the Thermoplasmatales^{9,14}. Group II euryarchaeotes are typically more abundant than crenarchaeotes in ocean surface waters^{9–11}, and blooms of these microorganisms have been observed in surface waters of Monterey Bay and the North Sea during summer^{13,14}. A few genome fragments from these planktonic euryarchaeotes have now been characterized^{19,20}, but little else is known about their genetic makeup, physiology or ecology.

Prokaryotic rhodopsins, first discovered in extremely halophilic Archaea (haloarchaea), are membrane proteins that bind retinal and respond to light stimuli. The known functional repertoire of these photoproteins now includes energy-conserving transmembrane proton pumps (bacteriorhodopsins, proteorhodopsins and xanthorhodopsins), transmembrane chloride pumps (halorhodopsins) and light sensors (sensory rhodopsins)^{1,21}. Among the haloarchaea, multiple rhodopsins with diversified functions can exist within a single cell. For example, the haloarchaeon *Haloarcula marismortui* genome encodes six rhodopsins: one proton-pumping

bacteriorhodopsin, one chloride-pumping halorhodopsin, two sensory rhodopsins, and two opsins of unknown function²² (Fig. 1a).

Prokaryotic rhodopsins were originally thought to exist exclusively in halophilic Archaea, but recent environmental genomic studies have revealed the existence, distribution and variability of a new class of such photoproteins, called proteorhodopsins, in members of the domain Bacteria^{1–7,23} (Fig. 1a). Biochemical and biophysical characterization of heterologously expressed proteorhodopsins indicates their potential functional role in light-activated proton translocation across bacterial membranes^{1,2,7}. Although planktonic Bacteria harbouring proteorhodopsins may not be photosynthetic *sensu stricto* with respect to carbon fixation, they probably gain a competitive advantage by using these photoproteins to generate a light-driven chemiosmotic potential. So far, proton-pumping proteorhodopsins have been clearly documented only in marine proteobacteria^{1–3,6}, including the recently cultivated and ubiquitous marine bacterium *Pelagibacter ubique*^{4,5}.

In a survey of the genomic diversity of marine picoplankton, we randomly sequenced both ends of large-insert DNA clones recovered from marine picoplankton in the North Pacific Subtropical Gyre²⁴. A fosmid clone (HF70_39H11) was identified that encoded a proteorhodopsin protein adjacent to a euryarchaeal-like LysE homologue on one end, and a euryarchaeal-like lysyl-tRNA synthetase on the other (Fig. 2). The presence of a bacterial-like proteorhodopsin on a planktonic archaeal genome fragment was unexpected, because no proteorhodopsin-like genes have previously been reported in Archaea. Two fosmid clones containing the novel proteorhodopsin (HF70_39H11 and HF70_59C08) were fully sequenced and compared (see Methods). The two clones formed a pseudocontig (Fig. 2) sharing 16 kilobase pairs of sequence overlap having 97% DNA sequence identity and complete gene synteny, except for one hypothetical gene unique to each (Supplementary Tables S1 and S2). Clone HF70_59C08 also contained genes encoding 22 ribosomal proteins, a preprotein translocase SecY subunit, and the SSU rRNA. Phylogenetic analyses of all these genes clearly identified both fosmids as deriving from marine group II euryarchaeotes (Fig. 3 and Supplementary Fig. S1). Heterologous expression of the HF70_59C08 proteorhodopsin in *Escherichia coli* confirmed that the archaeal photoprotein binds retinal covalently, and preliminary experiments indicate its potential function as a light-driven transmembrane proton pump *in vivo* (A.M., N.U.F. and E.F.D., unpublished observations).

We screened the large-insert DNA clone libraries prepared from various depths²⁴ to determine the evolutionary relatedness and environmental distribution of planktonic euryarchaeotes and archaeal proteorhodopsins in the water column. A total of 9,216 fosmid clones from each depth (330 Mb of cloned DNA per depth) were hybridized with a planktonic euryarchaeal SSU rRNA gene probe and an archaeal proteorhodopsin gene probe (see Methods).

¹Department of Civil and Environmental Engineering and Division of Biological Engineering, Massachusetts Institute of Technology, Building 48, 15 Vassar Street, Cambridge, Massachusetts 02139, USA. †Present address: Institute of Molecular Biology and Physiology, University of Copenhagen, Sølvgade 83H, 1307 Copenhagen K, Denmark.

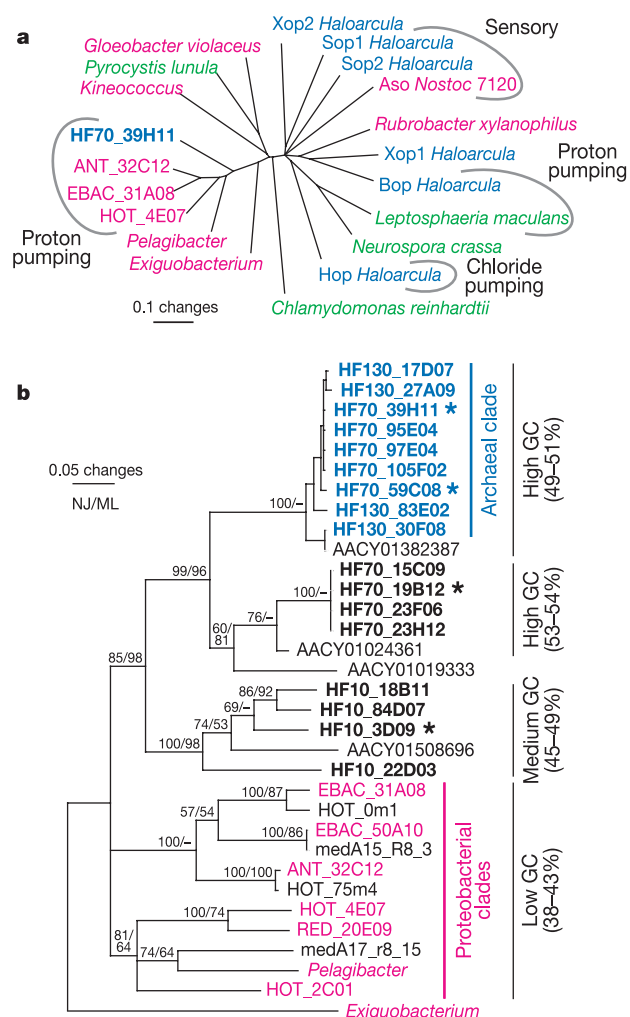


Figure 1 | Phylogenetic relationships of rhodopsins. **a**, Selected microbial rhodopsins representing their broad organismal and functional diversity. **b**, Proteorhodopsin-like rhodopsins (GC content refers to the nucleotide composition of the proteorhodopsin genes). Sequences obtained in this study are marked in bold. Fully sequenced fosmid clones are marked with an asterisk. Colours indicate organismal affiliation: blue, Archaea; magenta, Bacteria; green, Eukarya; black, uncertain. The trees are based on protein sequences and neighbour-joining distance calculations (see Methods). Bootstrap support (percentage) based on 1,000 replications is shown for neighbour-joining (NJ; first number) and maximum-likelihood (ML; second number) trees. Nodes with less than 50% support were collapsed.

Archaeal SSU rRNA genes identified in fosmid clones from each depth were sequenced to determine their phylogenetic affiliation. About 96% of all archaeal rRNA genes recovered in the upper water column were affiliated with group II euryarchaeotes (Fig. 3). Archaeal proteorhodopsins were also PCR amplified and sequenced, and compared with previously characterized rhodopsins (Fig. 1b and Supplementary Fig. S2).

Nine of the sequenced proteorhodopsins, including those from the fully sequenced fosmid clones, clearly originated from group II euryarchaeotes on the basis of the genetic linkage of the proteorhodopsin gene with archaeal SSU rRNA (Fig. 3). The archaeal proteorhodopsins formed a phylogenetic clade separate from other known proteobacterial proteorhodopsins (Fig. 1b) and had a G+C content of 49–51%, significantly higher than that of most other previously reported proteorhodopsin genes. Two other fully sequenced fosmids contained archaeal-like proteorhodopsins that fell into either the other high G+C clade (HF70_19B12), or the medium G+C clade (HF10_3D09) (Fig. 1b and Supplementary Fig. S3). The gene content

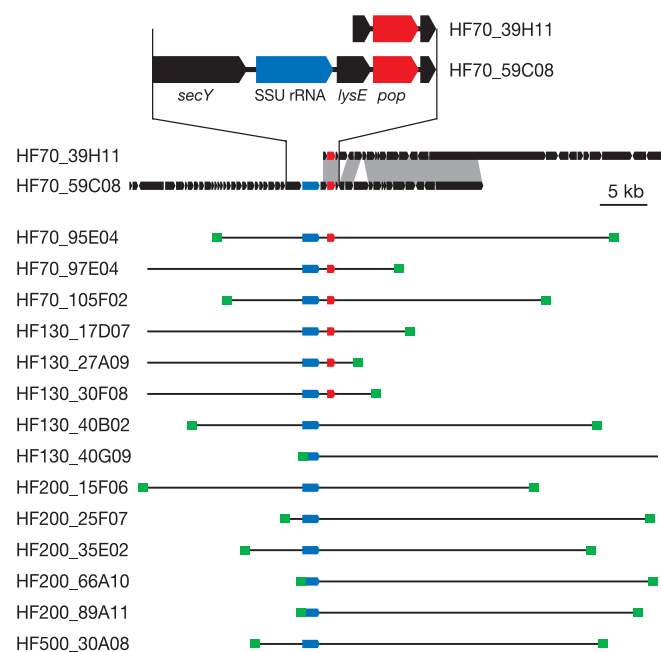


Figure 2 | Genetic maps and alignments of fosmid clones from the picoplankton genomic DNA libraries. Two fosmid clones containing archaeal SSU rRNA (HF70_39H11 and HF70_59C08) were fully sequenced. Proteorhodopsin, the SSU rRNA gene, and both clone termini were sequenced for the other fosmids shown. Colours indicate sequence type: blue, SSU rRNA gene; red, proteorhodopsin gene; green, clone terminus. Numbers directly following the HF clone designator indicate the depth of origin of the clone library: HF70, 70 m; HF130, 130 m; HF200, 200 m.

of these fosmids also suggested a euryarchaeal origin, and indicated that some archaeal-like proteorhodopsins are found in alternative genomic contexts unlinked to the SSU rRNA gene (Supplementary Fig. S3). Very few closely related homologues of the archaeal proteorhodopsins were found in the recently reported Sargasso Sea shotgun sequence data set²³, where 90% of all proteorhodopsin gene fragments screened had a G+C content of 35–45% (Supplementary Fig. S4), with most clustering in proteobacterial clades.

Fosmid clones encoding both euryarchaeal SSU rRNA and proteorhodopsin (Fig. 3) all shared near-identical SSU rRNAs (minimum 99.7% nucleotide sequence identity) and proteorhodopsins (minimum 95% amino acid sequence identity). In addition, terminal sequences of euryarchaeal rRNA-containing fosmids shared near-complete synteny with the two fully sequenced clones (HF70_39H11 and HF70_59C08). The coherent phylogenetic clustering of the linked SSU rRNA and proteorhodopsins on these genome fragments (Fig. 2), and the absence of their linkage in closely related, co-occurring strains (Figs 2 and 3), indicate that this arrangement might have been acquired only recently. Because all available data suggest that planktonic euryarchaeote genomes contain only a single SSU rRNA, this gene linkage seems to occur in about 10% of all the euryarchaeotes detected in the photic zone (8 of a total of 82 clones; Fig. 3). The remaining archaeal-like proteorhodopsin genes (unlinked to the SSU rRNA) seem to reside in different regions of the euryarchaeal genome, as indicated by photoproteins existing in different genetic contexts on euryarchaeal fosmids (HF70_19B12 and HF10_3D09; Fig. 1b) that lack the SSU rRNA gene (Supplementary Tables S3 and S4, and Supplementary Fig. S3).

Substantial numbers of euryarchaeal SSU rRNA genes and archaeal-type proteorhodopsin genes were observed in the photic zone, in roughly equal proportions (Fig. 4). In contrast, while euryarchaeal SSU rRNA genes were well represented in the sub-photic zone (200 m and below), archaeal-type proteorhodopsins were virtually absent (Fig. 4). Several archaeal clones lacked

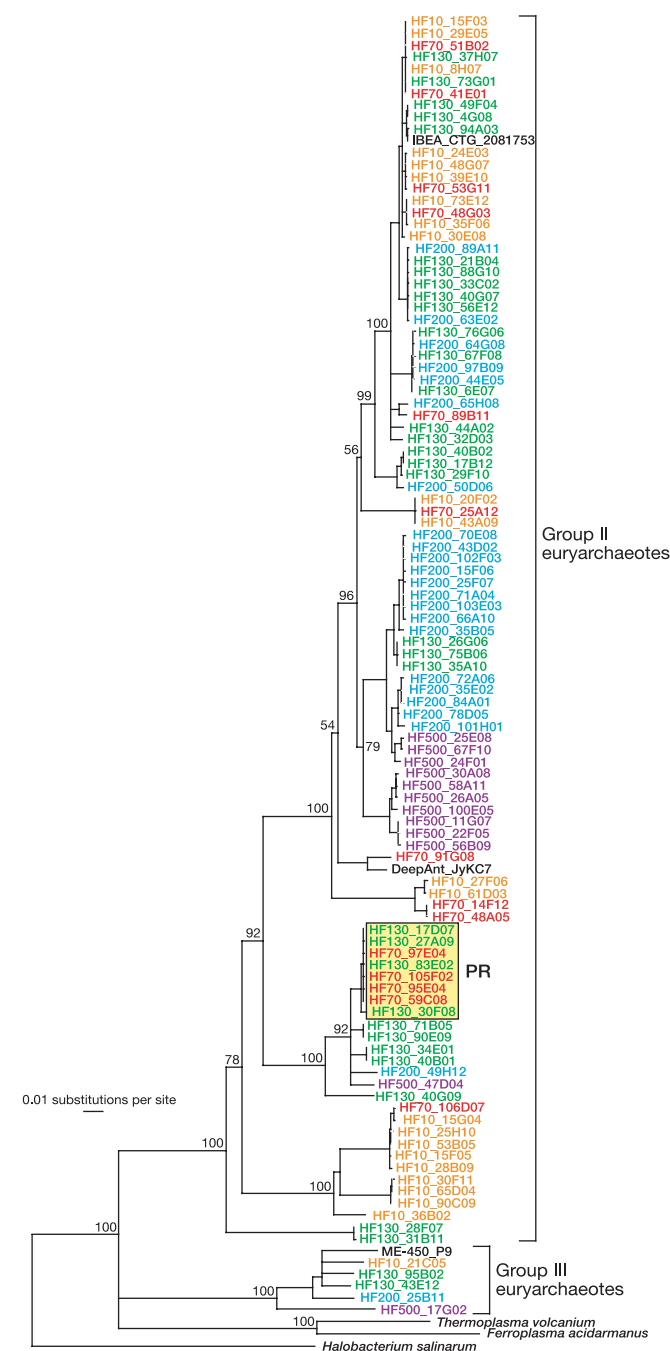


Figure 3 | Phylogenetic tree of euryarchaeal SSU rRNA sequences from the picoplankton genomic DNA libraries. Numbers directly following the HF clone designator indicate the depth of origin of each clone. Origins are also colour-coded: orange, 10 m; red, 70 m; green, 130 m; blue, 200 m; purple, 500 m; black, other origin. Clones containing both SSU rRNA and proteorhodopsin genes are indicated by PR. The tree is based on neighbour-joining distance calculations with Jukes-Cantor correction (bootstrap support (percentage) based on 1,000 replications is shown). *Methanosarcina maezei* was used as an outgroup (not shown). Nodes with less than 50% support were collapsed.

proteorhodopsin, but otherwise appeared syntenic with proteorhodopsin-containing clones based on SSU rRNA gene presence and aligned terminal sequences (Fig. 2). This group shares a minimum of 88% SSU rRNA sequence identity with the proteorhodopsin-containing types, indicating substantial genetic heterogeneity within group II euryarchaeotes. These phylotypes, derived primarily from the subphotic zone, represent lineages of group II euryarchaeotes

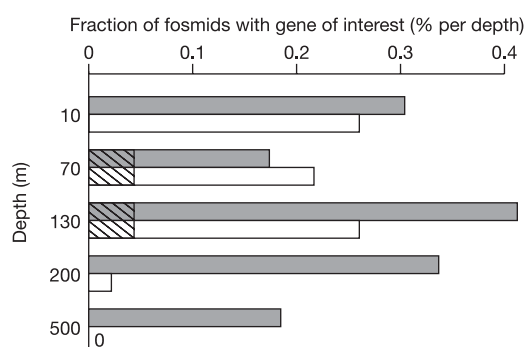


Figure 4 | Fosmid distribution in the picoplankton genomic DNA libraries. Bars indicate the fraction of fosmid clones at each depth interval that contain a euryarchaeal SSU rRNA gene (grey), an archaeal-like proteorhodopsin gene (white), or both a euryarchaeal SSU rRNA gene and an archaeal-like proteorhodopsin gene (hatched). $n = 9,216$ clones assayed at each depth.

related to the proteorhodopsin-containing genotypes but apparently lacking proteorhodopsin anywhere in their genomes (Fig. 4). Planktonic euryarchaeotes in the subphotic zone would presumably gain no competitive advantage from proteorhodopsin, so it is not surprising that they either have not acquired the photoprotein gene, or have not retained it.

Microbial rhodopsin-based photosystems are relatively simple and are encoded by very few genes. The retinal chromophore can be generated in one enzymatic step by the oxidative cleavage of a carotenoid^{7,25,26}. Assuming that carotenoid biosynthesis is present, just one gene enabling this enzymatic step, and another encoding an opsin, are sufficient to generate a functional rhodopsin. The genetic simplicity of these photosystems, their ability to assemble and function properly in the membranes of divergent microbial groups, and their potential to contribute to cellular energy metabolism, all are indicative of their likely predisposition for genetic mobility. Apparently, lateral gene dispersal mechanisms, coupled with strong selection for proteorhodopsin in the light, have contributed to the distribution of these photoproteins among various members of all three of life's domains. Proteorhodopsins therefore represent a category of 'cosmopolitan genes'²⁷ whose broad phylogenetic distribution is driven in part by lateral gene transfer, which further influences the recipient lineage's evolution and speciation^{28,29}. The spatial distribution of such promiscuous genes, including those encoding proteorhodopsin, appears more reflective of their functional properties in relation to the environment than of the specific organisms that harbour them.

METHODS

Seawater samples collected at 22°45' N, 158°00' W (100 km north of Oahu, Hawaii) in October 2002 were passed through a 1.6- μ m prefilter and retained on a 0.22- μ m filter²⁴. Large-insert genomic DNA fosmid clone libraries were prepared by using the vector pCC1FOS (Epicentre) as described previously³⁰. Clones were spotted on Hybond-N⁺ nylon filters (Amersham Biosciences) by Amplicon Express (Pullman, Washington). Hybridization was performed at 60°C with polymerase chain reaction (PCR)-generated DNA probes with the use of AlkPhos Direct Labelling and ECF Chemifluorescent Detection kits (Amersham Biosciences). PCR amplification of proteorhodopsin genes was done either with primer pair ArPRfor1 (5'-GACTATGTGGGTATTTCC-3') and ArPRrev1 (5'-GCCGAATGCGGTCTTATTGACCAAAATC-3') or with primer pair o-PR2 (5'-WWNMGNTAYGTNGAYTGG-3') and o-PR3 (5'-GGRTA DATNGCCANCC-3')⁷. PCR amplification of SSU rRNA genes was performed with primers Ar20F (5'-TTCCGGTTGATCCYGCC-3') and U1390R (5'-GACGGCGGTGTGTRC-3'). PCR products were sequenced directly with an ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). PCR products made with degenerate primers were cloned with the TOPO TA Cloning for Sequencing kit (Invitrogen) before being sequenced. Fosmid clones were fully sequenced with the TOPO Shotgun Subcloning kit (Invitrogen)

combined with sequence assembly with Sequencher v. 4.5 (Gene Codes Corporation) and annotation with FGENESB (Softberry) and Artemis v. 6 (The Wellcome Trust Sanger Institute). Sequences were aligned with the use of ClustalX v. 1.83 ([ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/](http://ftp-igbmc.u-strasbg.fr/pub/ClustalX/)) and SeaView (<http://pbil.univ-lyon1.fr/software/seaview.html>), and phylogenetic analyses were performed with PAUP 4.0 (Sinauer Associates) and Tree-Puzzle v. 5.2 (<http://www.tree-puzzle.de/>).

Received 8 August; accepted 15 November 2005.

- Béjà, O. *et al.* Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
- Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
- de la Torre, J. R. *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA* **100**, 12830–12835 (2003).
- Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
- Giovannoni, S. J. *et al.* Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**, 82–85 (2005).
- Sabehi, G., Béjà, O., Suzuki, M. T., Preston, C. M. & DeLong, E. F. Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.* **6**, 903–910 (2004).
- Sabehi, G. *et al.* New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol.* **3**, 1–9 (2005).
- Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaeobacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
- DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
- Massana, R., Murray, A. E., Preston, C. M. & DeLong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Appl. Environ. Microbiol.* **63**, 50–56 (1997).
- Massana, R., DeLong, E. F. & Pedrós-Alió, C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* **66**, 1777–1787 (2000).
- Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
- Pernthaler, A., Preston, C. M., Pernthaler, J., DeLong, E. F. & Amann, R. Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine Bacteria and Archaea. *Appl. Environ. Microbiol.* **68**, 661–667 (2002).
- DeLong, E. F. Oceans of *Archaea*. *ASM News* **69**, 503–511 (2003).
- Pearson, A., McNichol, A. P., Benitez-Nelson, B. C., Hayes, J. M. & Eglinton, T. I. Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compound-specific $\Delta^{14}\text{C}$ analysis. *Geochim. Cosmochim. Acta* **65**, 3123–3137 (2001).
- Wuchter, C., Schouten, S., Boschker, H. T. & Sinninghe Damsté, J. S. Bicarbonate uptake by marine Crenarchaeota. *FEMS Microbiol. Lett.* **219**, 203–207 (2003).
- Herndl, G. J. *et al.* Contribution of *Archaea* to total prokaryotic production in the deep Atlantic Ocean. *Appl. Environ. Microbiol.* **71**, 2303–2309 (2005).
- Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
- Béjà, O. *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2**, 516–529 (2000).
- Moreira, D., Rodríguez-Valera, F. & López-García, P. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ. Microbiol.* **6**, 959–969 (2004).
- Balashov, S. P. *et al.* Xanthorhodopsin: A proton pump with a light-harvesting carotenoid antenna. *Science* **309**, 2061–2064 (2005).
- Baliga, N. S. *et al.* Genome sequence of *Haloarcula marismortui*: A halophilic archaeon from the Dead Sea. *Genome Res.* **14**, 2221–2234 (2004).
- Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
- DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* (in the press).
- Giuliano, G., Al-Babili, S. & von Lintig, J. Carotenoid oxygenases: Cleave it or leave it. *Trends Plant Sci.* **8**, 145–149 (2003).
- Ruch, S., Beyer, P., Ernst, H. & Al-Babili, S. Retinal biosynthesis in Eubacteria: *in vitro* characterization of a novel carotenoid oxygenase from *Synechocystis* sp. PCC 6803. *Mol. Microbiol.* **55**, 1015–1024 (2005).
- Woese, C. R. A new biology for a new century. *Microbiol. Mol. Biol. Rev.* **68**, 173–186 (2004).
- Ochman, H., Lerat, E. & Daubin, V. Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl Acad. Sci. USA* **102** (Suppl. 1), 6595–6599 (2005).
- Ochman, H., Lerat, E., Daubin, V. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130 (2005).
- Suzuki, M. T. *et al.* Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb. Ecol.* **48**, 473–488 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Preston and L. Christianson for assistance in the initial sample collection and fosmid library construction; P. Richardson, K. Barry, S. Pitluck and the Joint Genome Institute production team for their help in sequencing; and D. Karl at the University of Hawaii, and the Hawaii Ocean Time Series team, who made sample collection possible. This research was supported by a grant to E.F.D. from The Gordon and Betty Moore Foundation under the Marine Microbiology Initiative program, and a National Science Foundation Microbial Observatory grant to E.F.D. Terminus sequencing of large-insert clones was performed by the Joint Genome Institute (Walnut Creek, California) under the auspices of the US Department of Energy's Office of Science, Biological, and Environmental Research Microbial Genomes Program.

Author Contributions E.F.D. made the initial bioinformatic observation of linkage between proteorhodopsin and archaeal genes. N.-U.F. and E.F.D. developed the concepts of the paper together. N.-U.F. performed the experiments, except the proteorhodopsin expression experiments, which were performed by A.M. T.J.M. participated in obtaining and analysing the SSU rRNA sequences. N.-U.F. wrote the first draft of the paper, which was completed by N.-U.F. and E.F.D. together.

Author Information The sequences reported here have been deposited in GenBank under accession numbers DQ257435, DQ257434, DQ156349, DQ156348 (fosmids HF10_3D09, HF70_19B12, HF70_39H11 and HF70_59C08), DQ156379–DQ156483 (SSU rRNA), DQ156350–DQ156363, DU708536–DU708556 (fosmid termini) and DQ156364–DQ156378 (proteorhodopsin sequences). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.F.D. (delong@mit.edu).

Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans

Timothy J. Aitman¹, Rong Dong^{1*}, Timothy J. Vyse^{2*}, Penny J. Norsworthy^{1*}, Michelle D. Johnson¹, Jennifer Smith³, Jonathan Mangion¹, Cheri Robertson-Lowe^{1,2}, Amy J. Marshall¹, Enrico Petretto¹, Matthew D. Hodges¹, Gurjeet Bhargal³, Sheetal G. Patel¹, Kelly Sheehan-Rooney¹, Mark Duda^{1,3}, Paul R. Cook^{1,3}, David J. Evans³, Jan Domin³, Jonathan Flint⁴, Joseph J. Boyle⁵, Charles D. Pusey³ & H. Terence Cook⁵

Identification of the genes underlying complex phenotypes and the definition of the evolutionary forces that have shaped eukaryotic genomes are among the current challenges in molecular genetics^{1–3}. Variation in gene copy number is increasingly recognized as a source of inter-individual differences in genome sequence and has been proposed as a driving force for genome evolution and phenotypic variation^{3–5}. Here we show that copy number variation of the orthologous rat and human *Fcgr3* genes is a determinant of susceptibility to immunologically mediated glomerulonephritis. Positional cloning identified loss of the newly described, rat-specific *Fcgr3* paralogue, *Fcgr3-related sequence* (*Fcgr3-rs*), as a determinant of macrophage overactivity and glomerulonephritis in Wistar Kyoto rats. In humans, low copy number of *FCGR3B*, an orthologue of rat *Fcgr3*, was associated with glomerulonephritis in the autoimmune disease systemic lupus erythematosus. The finding that gene copy number polymorphism predisposes to immunologically mediated renal disease in two mammalian species provides direct evidence for the importance of genome plasticity in the evolution of genetically complex phenotypes, including susceptibility to common human disease.

Glomerulonephritis is a major cause of kidney failure in humans, and is one of the most serious complications of autoimmune disorders such as systemic lupus erythematosus (SLE). In its most severe form, necrosis of mesangial and endothelial cells with accumulation of inflammatory and epithelial cells in Bowman's space give rise to the morphological appearance of glomerular 'crescents' or crescentic glomerulonephritis. Several rodent models of crescentic glomerulonephritis have been developed. The Wistar Kyoto (WKY) rat strain is uniquely susceptible to crescentic glomerulonephritis among rat strains tested, as demonstrated by susceptibility to experimentally induced nephrotoxic nephritis (NTN) (ref. 6 and H. Rennke, personal communication) and experimental autoimmune glomerulonephritis⁷. Glomerular injury and crescent formation in rat glomerulonephritis are macrophage-dependent⁸, with morphological changes closely resembling those seen in human focal and segmental necrotizing glomerulonephritis⁹. The genetic complexity of glomerulonephritis in humans led us to study the genetics of glomerulonephritis in WKY rats.

Ten days after injection of nephrotoxic serum, WKY rats consistently developed NTN, whereas Lewis, Brown Norway and Wistar rats showed no crescent formation, minimal glomerular macrophage infiltration and no proteinuria (Supplementary Fig. 1 and data not shown). F₁ rats showed intermediate phenotypes for crescent formation, proteinuria and macrophage infiltration, and phenotypes in

F₂ rats spanned the range of the parental strains (Supplementary Fig. 1). Crescent formation in F₂ rats was highly correlated with proteinuria ($r^2 = 0.66$; $P < 0.0001$) and with macrophage infiltration ($r^2 = 0.30$; $P < 0.0001$). Heritability was 0.96 for crescent formation, 0.80 for proteinuria and 0.52 for macrophage infiltration.

A genome screen for NTN susceptibility loci in F₂ rats revealed two major quantitative trait loci (QTLs) on chromosomes 13 and 16 (designated crescentic glomerulonephritis 1 (*Crng1*) and 2 (*Crng2*)), both of which were linked to crescent formation and proteinuria (logarithm of the odds (LOD) scores 7.4–9.1; Fig. 1a). Infiltration of macrophages was also strongly linked only to *Crng1* (LOD 6.1; Fig. 1b, c). Several additional linkages (LOD scores 3–4) to crescent formation and proteinuria were detected on other chromosomes and were designated *Crng3*–7. *Crng1* and *Crng2* accounted, respectively, for 21.8% and 16.8% of the genetic variance in crescent formation and 16.7% and 17.7% of the genetic variance in proteinuria. *Crng1* accounted for 12.9% of the genetic variance in glomerular macrophage infiltration.

Several biological candidates were found in the *Crng1* region of linkage including the genes encoding the activatory Fc receptor for IgG, *Fcgr3* (also known as FcγRIII), the inhibitory Fc receptor *Fcgr2* (FcγRII), and the common γ-subunit *Fcrlg* (FcRγ). We sequenced the coding region of *Fcrlg* and found no sequence variants. In *Fcgr2* we found a single nucleotide substitution, G388T, which changed codon 103 from arginine in Lewis rats to leucine in WKY rats. Sequence analysis of *Fcgr3* revealed evidence of a genomic rearrangement involving the *Fcgr3* locus and was therefore investigated further.

Polymerase chain reaction (PCR) amplification of *Fcgr3* exons from genomic DNA revealed single bands for exons 1–4 in Lewis and WKY rats, and a single band from WKY exon 5 but two discrete PCR products from Lewis exon 5 (Fig. 1d), suggesting duplication of exon 5 in Lewis genomic DNA, with loss of the shorter exon 5 from WKY genomic DNA. Direct sequence analysis of gel-purified products of exon 5 genomic DNA indicated the presence of a 226-base pair (bp) sequence in the longer PCR product (designated exon5_226+) that was absent in the shorter product (exon5_226–). The 226-bp sequence was situated in the 3'-untranslated region and contained a 153-bp short interspersed repetitive element (SINE).

PCR analysis of 27 divergent rat strains showed an identical pattern to Lewis, whereas spontaneously hypertensive rats (SHR) and stroke-prone SHR (SHRSP) showed the same pattern as WKY. Because WKY, SHR and SHRSP rats were derived in the mid-twentieth century from the same outbred Kyoto colony of Wistar rats¹⁰, we determined the genotype of five additional strains derived

¹Physiological Genomics and Medicine Group, MRC Clinical Sciences Centre, and Sections of ²Rheumatology and ³Renal Medicine, Imperial College, London W12 0NN, UK.

⁴Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. ⁵Department of Histopathology, Imperial College, London W12 0NN, UK.

*These authors contributed equally to this work.

from this colony. As with WKY, four of these also show loss of exon5₂₂₆– (Fig. 1e). The loss of exon5₂₂₆– from seven strains, all descended from the Kyoto Wistar colony, suggests that these strains have inherited this chromosomal segment identical-by-descent.

Because mesangial cell damage and glomerular necrosis are key features of NTN and are macrophage-dependent, we developed an *in vitro* assay of macrophage-mediated killing of antibody-coated glomerular mesangial cells as an intermediate phenotype of glomerular pathology in NTN. Macrophages from WKY rats showed markedly enhanced antibody-dependent cellular cytotoxicity (ADCC) compared with macrophages from Lewis rats (Fig. 2a). We further phenotyped Lewis, WKY and five of the other Kyoto-derived strains using a semi-automated, fluorescence-based assay of ADCC and found that five strains (WKY, SHR, WTC, WKYO, DON) had increased macrophage killing compared with Lewis, whereas one (IS/KYO) had macrophage activity comparable to Lewis (Fig. 2b). All of the strains with increased macrophage activity have lost exon5₂₂₆– from their genome. Haplotype analysis on chromosome 13 defined a 27-kilobase (kb) region, containing only one annotated gene (*Fcgr3*), that co-segregated with increased macrophage activity across these rat strains (Fig. 2c; see also Supplementary Information), excluding other genes in this region, including *Fcgr2*, as a cause of increased macrophage activity in WKY rats.

In addition to identification of the exon5₂₂₆+/- variant,

sequence analysis of *Fcgr3* exon 5 revealed a single nucleotide deletion in the coding sequence at position 129 (Δ G129), found only in the shorter exon 5 (exon5₂₂₆–) (Fig. 3a). Reverse transcriptase PCR (RT-PCR) showed that the Δ G129-containing exon 5 is transcribed (Fig. 3b), and we designate the gene from which this transcription product is derived as *Fcgr3-rs*. No copies of the Δ G129 variant were found in 65 separate clone inserts amplified from WKY complementary DNA or genomic DNA, confirming loss of *Fcgr3-rs* from the WKY genome.

The Δ G129 variant results in a frameshift in the *Fcgr3-rs* coding sequence that predicts a novel cytoplasmic domain six amino acids longer than that encoded by *Fcgr3*. Sequence comparison showed that although *Fcgr3-rs* is highly similar to the other isoforms of *Fcgr3* across the extracellular and transmembrane domains, the *Fcgr3-rs* cytoplasmic domain has no homology to other known proteins.

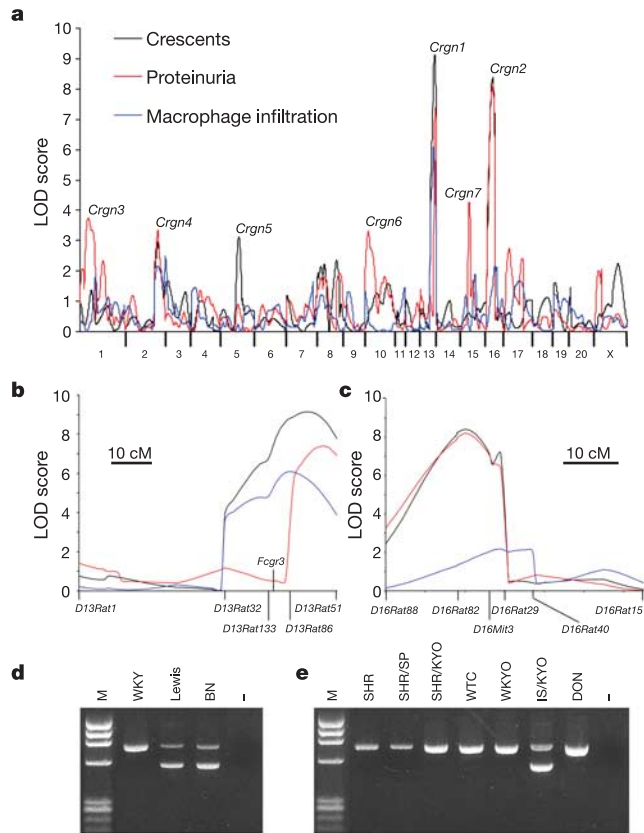


Figure 1 | Genome screen for NTN susceptibility loci, and duplication and loss of *Fcgr3* exon 5. **a–c**, Multi-point linkage plots showing location of susceptibility genes for crescent formation, proteinuria and macrophage infiltration on a whole genome plot (**a**), chromosome 13 (**b**) and chromosome 16 (**c**). cM, centimorgan. **d**, **e**, PCR amplification of *Fcgr3* exon 5 from Lewis and WKY rats with Brown Norway strain as reference, showing absence of exon5₂₂₆– in WKY (**d**), and in six out of seven other Kyoto-derived rat strains (**e**). –, negative control PCR; M, Φ X174 *Hae*III size marker.

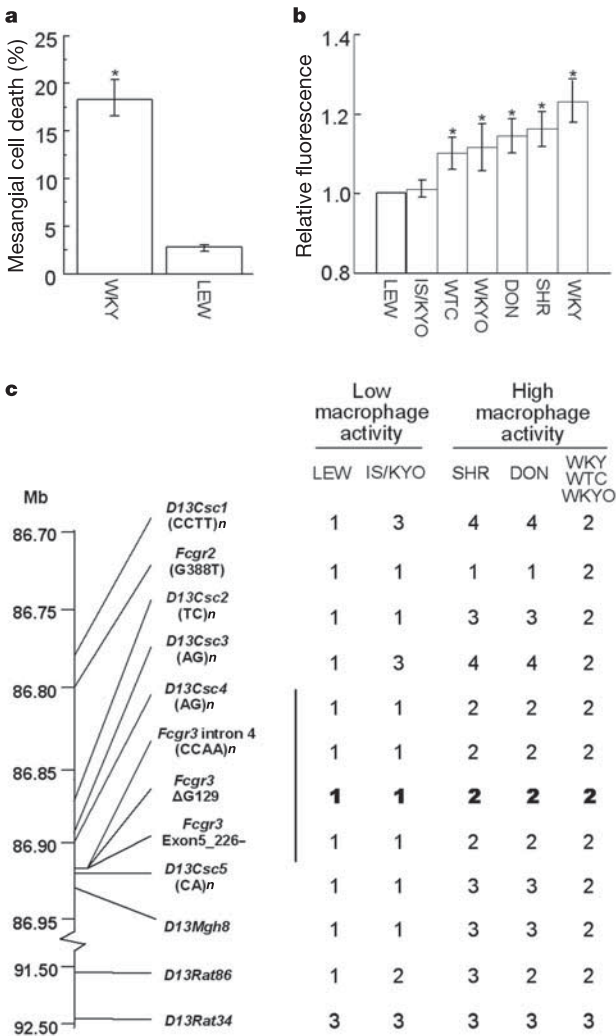


Figure 2 | Macrophage activity and haplotype analysis. **a**, **b**, Macrophage-mediated killing of antibody-coated mesangial cells by Lewis and WKY macrophages determined by visual assessment ($n = 5$ rats per strain; asterisk indicates $P < 0.001$) (**a**), and in different rat strains determined by release of fluorescence, normalized to Lewis macrophages ($n = 4$ rats per strain; asterisk indicates $P < 0.01$ compared to Lewis) (**b**). Data are mean $\pm$ s.e.m. **c**, Chromosome 13 haplotype. The vertical bar indicates the interval shared between rat strains with macrophage overactivity. No other marker genotypes co-segregated with macrophage overactivity within the *Crgn1* linkage region. High and low macrophage activity denote, respectively, strains with macrophage activity that is significantly increased, or not, compared to Lewis. Numbers denote classification of alleles: 1, Lewis; 2, WKY; 3, 4, other alleles. The *Fcgr3-rs* Δ G129 genotype is indicated in bold.

Southern analysis confirmed the presence of at least three copies of *Fcgr3*-like exon 5 sequences in Lewis and WKY genomic DNA and reaffirmed loss of *Fcgr3*-rs exon 5 from the WKY genome (Fig. 3c); clonotype analysis (Supplementary Fig. 2) indicated the presence in both Lewis and WKY rats of at least two distinct transcribed copies of exon5_ins226+. Additional Southern analysis of two *Fcgr3*-containing bacterial artificial chromosomes from the rat genome project (data not shown) showed identically sized restriction fragments to those shown in Fig. 3c, indicating that these genes reside in contiguous genomic DNA on chromosome 13.

We then carried out western analysis of macrophage lysates from Lewis and WKY rats using anti-*Fcgr3*-rs-specific antiserum. This detected a 66-kDa band in Lewis but not in WKY rats (Fig. 3d), confirming expression of *Fcgr3*-rs protein in Lewis macrophages and its deficiency in WKY.

To compare function of *Fcgr3* and *Fcgr3*-rs, we stably transfected COS-1 cells with the common γ -subunit (*Fcgr1g*) together with either the *Fcgr3* or *Fcgr3*-rs α -subunits. We also transfected COS-1 cells with the *Fcgr3* construct after modification by site-directed mutagenesis to delete the single nucleotide G129 from exon 5, yielding a chimaeric construct, *Fcgr3*- Δ G, that contains *Fcgr3* extra-cellular and transmembrane domains with an *Fcgr3*-rs cytoplasmic domain. We tested the various stably transfected COS-1 cells for their potential to phagocytose opsonized sheep red blood cells (SRBCs). Similar expression of all *Fcgr3* constructs was confirmed by western analysis using an antibody to the Myc tag (data not shown), and cell surface expression was confirmed by flow cytometry (Supplementary Fig. 3). Cells transfected with both *Fcgr1g* and *Fcgr3* showed levels of

phagocytosis over tenfold greater than cells transfected with *Fcgr1g* and either *Fcgr3*-rs or the *Fcgr3*- Δ G constructs (Fig. 3e). Co-transfection of *Fcgr3* with *Fcgr3*-rs showed a 70% inhibition of *Fcgr3*-mediated phagocytosis (Fig. 3f), indicating that loss of *Fcgr3*-rs-mediated inhibition of *Fcgr3* is the likely mechanism for macrophage over-activity in WKY compared to Lewis rats.

Our previous studies of Fc receptor polymorphisms in northern European nuclear families with SLE showed unexpected mendelian errors for two polymorphisms in the *FCGR3B* gene in 14% of these families¹¹. The association of *Fcgr3* copy number variation with immunologically mediated glomerulonephritis in the rat led us to test the hypothesis that copy number variation in human *FCGR3B* might explain these mendelian errors and that this could contribute to glomerulonephritis susceptibility.

We developed a quantitative PCR assay to measure *FCGR3B* gene copy number and applied this assay to 30 individuals from a subset of 8 of the nuclear families shown previously to have non-mendelian inheritance at *FCGR3B*. This showed significant variation in *FCGR3B* copy number that was consistent with an estimate by Southern analysis (Fig. 4a). *FCGR3B* copy number differed significantly in this sample from that expected for a single copy gene in a diploid genome ($P = 0.0004$; Supplementary Fig. 4).

We tested for an association between *FCGR3B* copy number and disease in SLE patients and in the subset of SLE patients with glomerulonephritis, referred to as lupus nephritis. Discordant sib pair analysis (one sib pair per family) showed no association between *FCGR3B* copy number and SLE ($n = 187$ sib pairs, Wilcoxon sign-rank test, $P_{2-tailed} = 0.083$, 95% confidence interval (CI) 0.082–0.085), but

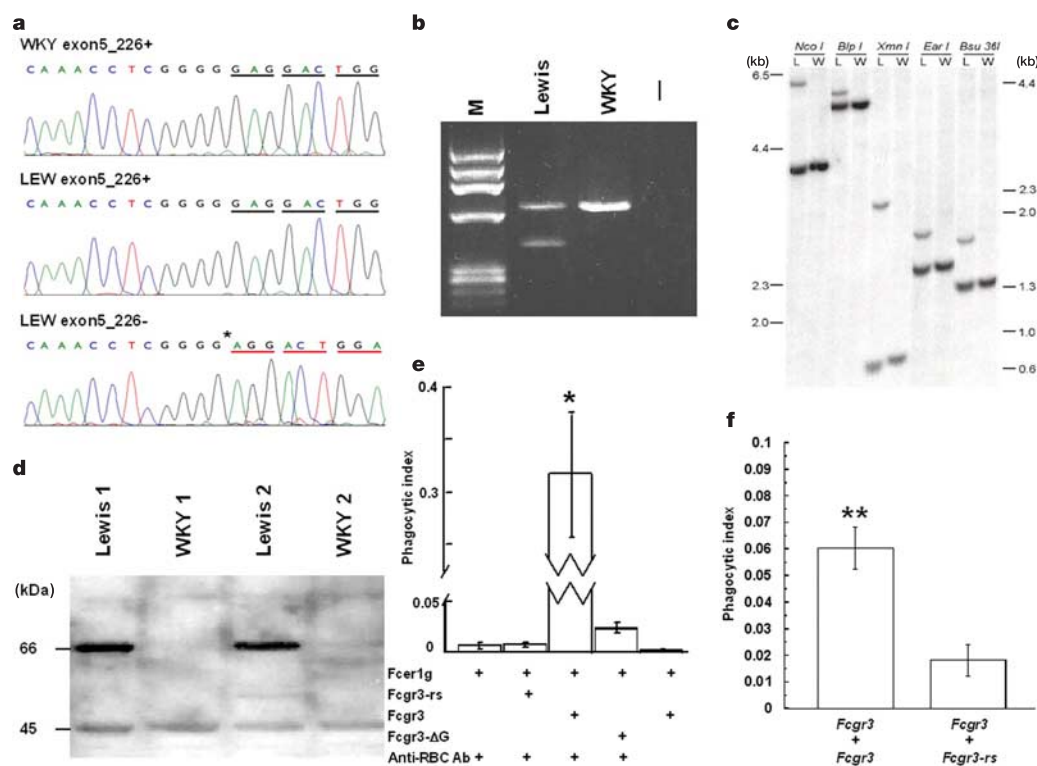


Figure 3 | Identification and functional characterization of *Fcgr3*-rs.

a, Sequence analysis of exon 5 in Lewis and WKY rats showing a guanine deletion at position 129 (Δ G129) in Lewis exon5_226-, denoted with an asterisk. The new reading frame is underlined in red. **b**, RT-PCR of Lewis and WKY RNA indicates expression of *Fcgr3*-rs exon5_226- in Lewis but not in WKY rats. M, Φ X174 *Hae*III size marker. **c**, Southern blot of the *Fcgr3* locus with genomic DNA from WKY (W) and Lewis (L) rats. All of the restriction enzymes cut within the exon 5 SINE-containing insertion, and do not cut within the probe sequence. *Xmn*I, *Ear*I and *Bsu*361 digests were

loaded after the *Nco*I and *Bsp*I digests and have separate size markers (kb). **d**, Western blot of peritoneal macrophage lysate using a custom antibody to the *Fcgr3*-rs cytoplasmic tail, showing presence of *Fcgr3*-rs in Lewis but not in WKY macrophages on replicate loadings. **e**, Phagocytosis of antibody-coated erythrocytes by COS-1 cells transfected with *Fcgr3* constructs. Asterisk, $P < 0.01$ compared to all other constructs. **f**, Co-transfection of *Fcgr3* with *Fcgr3*-rs showing inhibition of *Fcgr3*-mediated phagocytosis by *Fcgr3*-rs ($n = 8$ paired transfections). Double asterisk, $P < 0.002$. Data are mean $\pm$ s.e.m.

showed a weak association with lupus nephritis ($n = 61$ pairs, Wilcoxon sign-rank test, $P_{2\text{-tailed}} = 0.038$, 95% CI 0.037–0.039). Separate association analysis, carried out between all available lupus nephritis patients ($n = 63$) and unrelated, seronegative controls ($n = 141$) from the lupus cohort also showed association with lupus nephritis (Mann–Whitney U -test, $P = 0.001$, 95% CI 0.001–0.002; Fig. 4b). Logistic regression analysis, using age and gender as covariates, confirmed the association between *FCGR3B* copy number and lupus nephritis in this sample (χ^2 (3 degrees of freedom) = 13.5, $P = 0.004$; Supplementary Table 1A).

Because previous studies have shown an association between SLE and the *FCGR2A* and *FCGR3A* genes within the Fc receptor cluster on chromosome 1q23 (refs 12, 13), we carried out a sequential logistic regression analysis to test whether the effect of *FCGR3B* copy number on lupus nephritis was independent of *FCGR2A* and *FCGR3A*. Using data from 60 lupus nephritis patients and 109 unrelated, seronegative controls in this cohort for whom complete genotype data was available, the sequential logistic regression analysis model predicted *FCGR3B* copy number as being significantly and independently associated with lupus nephritis (χ^2 (3 degrees of freedom) = 13.4, $P = 0.004$), whereas the inclusion of *FCGR2A*-G548A and *FCGR3A*-T559G produced poorer models, indicating that the effects of *FCGR3A* and *FCGR2A* are very small compared with that observed for *FCGR3B* (Supplementary Table 1B). An analysis of *FCGR2A*-G548A and *FCGR3A*-T559G haplotypes in this group did not show any association with copy number at *FCGR3B* (Fisher's exact test, $P_{2\text{-tailed}} > 0.5$), providing further evidence that reduced copy number at *FCGR3B* is an independent risk factor for lupus nephritis.

Fc receptors, the genes of which are located in clusters across mammalian genomes, functionally link the humoral and cellular branches of the immune system and have a key role in activation and modulation of the immune response^{14,15}. Our findings in rats that loss of *Fcgr3*-rs results in macrophage overactivity and glomerulonephritis susceptibility, and that *Fcgr3*-rs inhibits *Fcgr3*-mediated phagocytosis, suggest that *Fcgr3*-rs may be a potential therapeutic agent for autoimmune disease. Human *FCGR3B* is expressed mainly in neutrophils and is necessary for neutrophil tethering to immune complexes^{15–17}. It is therefore plausible that reduced neutrophil expression of *FCGR3B* in patients with low *FCGR3B* copy number may lead to reduced glomerular clearance of immune complexes and susceptibility to autoimmune renal disease in patients with SLE, and possibly to other autoimmune disorders.

So far, there is little evidence that common, complex disease

phenotypes can be caused by stably transmitted gene duplication/deletion or copy number polymorphisms⁴. Two examples that have been reported are *Cd36* deletion in insulin resistance in rats and *CCL3L1* copy number polymorphism in HIV-1/AIDS susceptibility in humans^{18–20}. Although multiple genomic copies of *Fcgr3* have been suggested in the rat²¹, and gene duplication and deletion in human *FCGR3B* have been reported in isolated cases or in families identified because of *FCGR3B* deficiency, their association with immune phenotypes has not been clear^{22–24}. Our studies have demonstrated naturally occurring gene copy number variation within the syntenic *Fcgr3* clusters in rats and humans, and showed that loss of *Fcgr3*-rs in the rat and low *FCGR3B* copy number in humans are associated with susceptibility to immunologically mediated forms of glomerulonephritis in these two mammalian species. To our knowledge, this is the first demonstration in any species that gene copy number polymorphism predisposes to autoimmune disease. Our finding that copy number polymorphism at orthologous regions of diverse genomes is associated with immunologically related disease suggests that genome plasticity, manifested by gene duplication/deletion and copy number polymorphism, is a more common cause of genetically complex phenotypes than has hitherto been observed.

METHODS

See Supplementary Information for detailed Methods.

Inbred strains and linkage studies. WKY/NCrIbR (abbreviated to WKY) and SHR/NCrIbR (SHR) rats were obtained from Charles River and Lew/HanHsd (Lewis) rats from Harlan. WTC, WKYO, DON and IS/KYO rats were obtained from T. Serikawa. Reciprocal crosses between WKY and Lewis (both *Rt1-l*) rats were used to generate 177 F₂ rats that were phenotyped for NTN and used in a genome screen for susceptibility genes with 128 polymorphic microsatellite markers. No phenotypic differences were found between the reciprocal crosses, which were combined for linkage studies. NTN phenotypes were assessed in 200–220 g male rats 10 days after intravenous injection of 0.1 ml of a rabbit antiserum to rat glomerular basement membrane²⁵. Crescent formation, 24 h urinary protein and glomerular macrophage infiltration were assessed as described²⁶.

Direct sequencing and sequencing of cloned PCR products. RNA extraction from frozen tissues, cDNA preparation and PCR were undertaken as described¹⁸, and clonotype analysis was performed using the high-fidelity polymerases *Pfu* (Promega) or *KOD1* (Novagen).

Southern and western analyses. Southern analyses were carried out as described¹⁸. For western analysis, 10 μ g denatured protein from lysed thioglycollate-elicited macrophages or COS-1 cells was resolved by 4–12% gradient SDS–PAGE. Proteins were electro-blotted onto PVDF membranes (Invitrogen) and stained with antibodies against the *Fcgr3*-rs cytoplasmic domain (macrophages) or against Myc tag (COS-1 cells).

Antibody-directed cellular cytotoxicity. ADCC assays were carried out by visual assessment of cell death of antibody-coated mesangial cells and by a semi-automated dye release assay, as described²⁷. Mesangial cells were derived from WKY kidneys. Similar results were found with Lewis mesangial cells. For the dye release assay, thioglycollate-elicited peritoneal macrophages were added to cultured mesangial cells loaded with calcein fluorescent dye in the presence of 10 μ g ml⁻¹ anti-Thy1.1 antibody. Calcein release into the supernatant was measured in a microplate fluorimeter. Differences between inbred strains were assessed by Student's t -test (1-tailed).

Studies of Fc receptor function in COS-1 cells. Constructs of *Fcer1g*, *Fcgr3*, *Fcgr3*-rs and *Fcgr3*- Δ G were created from PCR-amplified cDNA. A Myc tag was introduced into the amino terminus of all three *Fcgr3* α -subunit constructs to allow confirmation of expression by western blot. COS-1 cells stably transfected with *Fcer1g* were electroporated with *Fcgr3*, *Fcgr3*-rs or *Fcgr3*- Δ G and selected with G418 and hygromycin B. Fc-receptor-mediated function was assessed by measuring internalization into COS-1 cells of SRBCs coated with anti-SRBC antibody. Phagocytic index (RBCs per COS-1 cell) was calculated as described²⁸. The effect of *Fcgr3*-rs on *Fcgr3*-mediated phagocytosis was determined by transfection of COS-1 cells that had been transfected with, and were expressing, *Fcer1g* and *Fcgr3*. Differences between groups were compared using Student's t -test (1-tailed).

Nucleotide and protein sequence analysis. Homologous sequences to rat *Fcgr3* and *Fcgr3*-rs exon 5 sequences were sought by searching the NCBI non-redundant and high-throughput genomic sequence databases using BLAST.

Quantification of *FCGR3B* copy number. Quantitative PCR was carried out

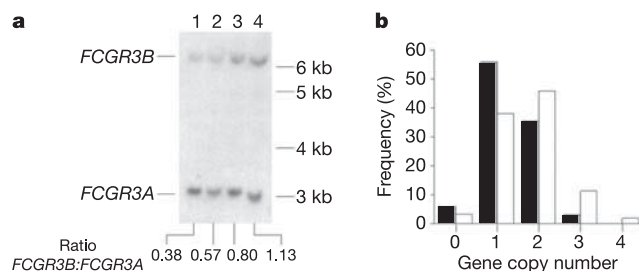


Figure 4 | Copy number polymorphism in human *FCGR3* and association with lupus nephritis. **a**, Southern blot of genomic DNA from four unrelated individuals, selected according to the quantitative PCR estimate of *FCGR3B* copy number, shown above the blot. The Southern probe was designed to cross-hybridize to *FCGR3A* and *FCGR3B* restriction fragments. *FCGR3B* band intensity, measured by densitometry, is normalized to *FCGR3A* band intensity, and shown below the plot as the ratio of *FCGR3B*:*FCGR3A*. Genomic DNA was digested with the restriction enzyme *HpaI*. **b**, Histogram showing frequency distribution of *FCGR3B* copy number, per diploid genome, in lupus nephritis cases (filled columns) and unaffected controls (open columns).

using SYBR Green Jumpstart *Taq* Readymix (Sigma) and analysed by the standard curve method. Oligonucleotides were designed to amplify specifically *FCGR3B* and to avoid paralogous or allelic sequence variants. *CD36* was used as a single-copy control. Data from *FCGR3B* and *CD36* were normalized to forkhead box P2 (*FOXP2*) to give an estimate of copy number. Sequence analysis was undertaken to confirm specificity of the *FCGR3B* PCR product.

Human genetics. DNA was available from 256 SLE nuclear families (consisting of parents and progeny) of northern European origin in which genotypes for *FCGR3B* had been previously determined²⁹. SLE and lupus nephritis were defined according to ACR criteria³⁰. This cohort consisted of a single SLE patient per family. Of the 256 SLE patients studied, 63 had nephritis and were available for analysis. Discordant sib pair analyses were performed for all families with available sib pairs (one sib pair per family) using the nonparametric Wilcoxon sign-rank test. We also compared the 63 lupus nephritis cases with unrelated controls ($n = 141$) (one per family) from within the SLE cohort. All controls used were seronegative for antinuclear antibody. The study was ethically approved under MREC 98/2/6.

A one-sample sign test of the median was used to determine whether gene copy number differed significantly from 2. To test for the significance of the difference between the distributions of the *FCGR3B* gene copy numbers we used the nonparametric Mann–Whitney *U*-test (2-tailed). For both the Wilcoxon sign-rank test and the Mann–Whitney *U*-test, 100,000 Monte Carlo simulations were performed to estimate exact *P* values and 95% confidence intervals. Logistic regression was performed using SPSS 12.0 with *FCGR3B* gene copy number and genotypes of *FCGR2A* and *FCGR3A* as predictor variables. Haplotypes for *FCGR2A-G548A* and *FCGR3A-T559G* were constructed as described¹¹.

Received 3 August; accepted 22 November 2005.

- Ohno, S., Wolf, U. & Atkin, N. B. Evolution from fish to mammals by gene duplication. *Heredity* **59**, 169–187 (1968).
- Glazier, A. M., Nadeau, J. H. & Aitman, T. J. Finding genes that underlie complex traits. *Science* **298**, 2345–2349 (2002).
- Eichler, E. E. & Sankoff, D. Structural dynamics of eukaryotic chromosome evolution. *Science* **301**, 793–797 (2003).
- Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
- Sebat, J. et al. Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- Kawasaki, K., Yaoita, E., Yamamoto, T. & Kihara, I. Depletion of CD8 positive cells in nephrotoxic serum nephritis of WKY rats. *Kidney Int.* **41**, 1517–1526 (1992).
- Sado, Y., Naito, I., Akita, M. & Okigaki, T. Strain specific responses of inbred rats on the severity of experimental autoimmune glomerulonephritis. *J. Clin. Lab. Immunol.* **19**, 193–199 (1986).
- Isome, M. et al. Important role for macrophages in induction of crescentic anti-GBM glomerulonephritis in WKY rats. *Nephrol. Dial. Transplant.* **19**, 2997–3004 (2004).
- Ferrario, F. & Pia Rastaldi, M. in *Rapidly Progressive Glomerulonephritis* (eds Pusey, C. D. & Rees, A. J.) 59–107 (Oxford Univ. Press, Oxford, 1998).
- Yamori, Y. in *Handbook of Hypertension Vol. 4 Experimental and Genetic Models of Hypertension* (ed. de Jong, W.) 224–239 (Elsevier, New York, 1984).
- Robertson, C. A. *Fine Mapping Candidate Gene Families in Human Systemic Lupus Erythematosus*. Thesis, Univ. London (2003).
- Karassa, F. B., Trikalinos, T. A. & Ioannidis, J. P. Role of the Fcγ receptor IIa polymorphism in susceptibility to systemic lupus erythematosus and lupus nephritis: a meta-analysis. *Arthritis Rheum.* **46**, 1563–1571 (2002).
- Edberg, J. C. et al. Genetic linkage and association of Fcγ receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. *Arthritis Rheum.* **46**, 2132–2140 (2002).
- Ravetch, J. V. Fc receptors: rubor redux. *Cell* **78**, 553–560 (1994).
- Takai, T. Roles of Fc receptors in autoimmunity. *Nature Rev. Immunol.* **2**, 580–592 (2002).
- Kurosaki, T. & Ravetch, J. V. A single amino acid in the glycosyl phosphatidylinositol attachment domain determines the membrane topology of FcγRIII. *Nature* **342**, 805–807 (1989).
- Coxon, A. et al. FcγRIII mediates neutrophil recruitment to immune complexes. A mechanism for neutrophil accumulation in immune-mediated inflammation. *Immunity* **14**, 693–704 (2001).
- Aitman, T. J. et al. Identification of *Cd36* (*Fat*) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21**, 76–83 (1999).
- Pravenec, M. et al. Transgenic rescue of defective *Cd36* ameliorates insulin resistance in spontaneously hypertensive rats. *Nature Genet.* **27**, 156–158 (2001).
- Gonzalez, E. et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
- Farber, D. L. et al. Rat class III Fcγ receptor isoforms differ in IgG subclass-binding specificity and fail to associate productively with rat CD3ζ. *J. Immunol.* **150**, 4364–4375 (1993).
- Clark, M. R., Liu, L., Clarkson, S. B., Ory, P. A. & Goldstein, I. M. An abnormality of the gene that encodes neutrophil Fc receptor III in a patient with systemic lupus erythematosus. *J. Clin. Invest.* **86**, 341–346 (1990).
- Koene, H. R., Kleijer, M., Roos, D., de Haas, M. & Von dem Borne, A. E. FcγRIIIB gene duplication: evidence for presence and expression of three distinct FcγRIIIB genes in NA(1+,2+)SH(+) individuals. *Blood* **91**, 673–679 (1998).
- Gittinger, F. S., Schindler-Wuepper, L., Kissel, K. & Bux, J. Quantitative determination of Fcγ receptor genes by means of fluorescence-based real-time polymerase chain reaction. *Tissue Antigens* **60**, 64–70 (2002).
- Bhan, A. K., Schneeberger, E. E., Collins, A. B. & McCluskey, R. T. Evidence for a pathogenic role of a cell-mediated immune mechanism in experimental glomerulonephritis. *J. Exp. Med.* **148**, 246–260 (1978).
- Cook, H. T. et al. Interleukin-4 ameliorates crescentic glomerulonephritis in Wistar Kyoto rats. *Kidney Int.* **55**, 1319–1326 (1999).
- Boyle, J. J. Human macrophages kill human mesangial cells by Fas-L-induced apoptosis when triggered by antibody via CD16. *Clin. Exp. Immunol.* **137**, 529–537 (2004).
- Kim, M. K. et al. Fcγ receptor transmembrane domains: role in cell surface expression, gamma chain interaction, and phagocytosis. *Blood* **101**, 4479–4484 (2003).
- Russell, A. I. et al. Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Hum. Mol. Genet.* **13**, 137–147 (2004).
- Tan, E. M. et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**, 1271–1277 (1982).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge intramural funding from the CSC, and support from the Wellcome Trust Cardiovascular Functional Genomics award, the British Heart Foundation and the Medical Research Council. P.R.C. is a Medical Research Council Clinical Fellow. We thank E. Sodergren and G. Weinstock for BAC clone DNA; H. Hedrich, A. Dominiczak and J. Rapp for rat genomic DNA; T. Serikawa and the National Bio Resource Project in Japan for rat strains and genomic DNA; B. Foxwell for bicistronic vector; and M. Botto, B. Morley, P. Froguel, C. Shoulders and S. Cook for constructive criticism of the manuscript. We acknowledge the CSC Genomics Laboratory for DNA sequencing, and bioinformatics support from M. Müller, N. J. Dickens and the Imperial College Bioinformatics Support Service.

Author Contributions The study was conceived and funded by T.J.A., H.T.C. and C.D.P. H.T.C., J.S., P.R.C. and D.J.E. carried out the rodent phenotyping. T.J.A., M.D., P.J.N., P.R.C. and J.F. carried out the rodent linkage studies. T.J.A., R.D., M.D.J., J.M., A.J.M., M.D.H., S.G.P. and K.S.-R. carried out the genomic analysis of rat and human *Fcgr3*. Cellular immunology studies were carried out by R.D., J.J.B., M.D.J., G.B., M.D., J.D., C.D.P. and H.T.C. Human genetics was carried out by P.J.N., A.J.M., C.R.-L., T.J.V., E.P. and T.J.A., and the manuscript was written by T.J.A., H.T.C., T.J.V. and J.M.

Author Information The sequence of *Fcgr3-rs* exon 5 has been deposited in GenBank under accession number AY561710. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J.A. (t.aitman@csc.mrc.ac.uk) or H.T.C. (t.h.cook@imperial.ac.uk).

LETTERS

A bottom-up approach to gene regulation

Nicholas J. Guido^{1*}, Xiao Wang^{2*}, David Adalsteinsson³, David McMillen⁵, Jeff Hasty⁶, Charles R. Cantor¹, Timothy C. Elston⁴ & J. J. Collins¹

The ability to construct synthetic gene networks enables experimental investigations of deliberately simplified systems that can be compared to qualitative and quantitative models^{1–23}. If simple, well-characterized modules can be coupled together into more complex networks with behaviour that can be predicted from that of the individual components, we may begin to build an understanding of cellular regulatory processes from the ‘bottom up’. Here we have engineered a promoter to allow simultaneous repression and activation of gene expression in *Escherichia coli*. We studied its behaviour in synthetic gene networks under increasingly complex conditions: unregulated, repressed, activated, and simultaneously repressed and activated. We develop a stochastic model that quantitatively captures the means and distributions of the expression from the engineered promoter of this modular system, and show that the model can be extended and used to accurately predict the *in vivo* behaviour of the network when it is expanded to include positive feedback. The model also reveals the counterintuitive prediction that noise in protein expression levels can increase upon arrest of cell growth and division, which we confirm experimentally. This work shows that the properties of regulatory subsystems can be used to predict the behaviour of larger, more complex regulatory networks, and that this bottom-up approach can provide insights into gene regulation.

We engineered the $O_R O_{lac}$ promoter in *E. coli* as a tool for investigating the interactive effects of positive and negative transcriptional regulation. The basis for the engineered promoter is the P_{RM} promoter of bacteriophage λ . In the native promoter, the operator sites O_R1 and O_R2 cooperatively bind the λ repressor protein, CI, which in turn acts as a transcriptional activator by recruiting RNA polymerase²⁴. In the engineered $O_R O_{lac}$ promoter, the third operator site of the native P_{RM} promoter, O_R3 , was mutated to significantly reduce CI binding, thereby removing the repressive function of this site. A *lac* operator, which can bind the repressor protein LacI, was placed upstream of O_R1 . The result is a promoter that can be repressed by LacI and activated by CI; it thus enables the combined effects of repression and activation to be investigated.

We inserted the $O_R O_{lac}$ promoter in a high-copy plasmid that lacked the *cl* activator and *lacI* repressor genes to create an unregulated system (Fig. 1a), to establish the promoter’s basal properties. The green fluorescent protein gene (*gfp*) was placed under the control of the $O_R O_{lac}$ promoter as a readout for transcriptional induction. A repressor-only system was constructed by adding a strongly constitutively expressed *lacI* gene, under the control of the P_{LtetO1} promoter, to the unregulated system (Fig. 1b). The LacI protein binds to the *lac* operator as a tetramer, which reduces the binding of the RNA polymerase, and may affect the binding of CI to its operator

site. When isopropyl- β -D-thiogalactopyranoside (IPTG) is added to the media, it binds to the LacI tetramer and weakens binding of the LacI protein to the O_{lac} operator site. This allowed us to tune the degree of $O_R O_{lac}$ repression. An activator-only system was constructed by adding the *cl* gene under the control of the *pBAD* promoter, which is activated by arabinose (Fig. 1c). The CI protein forms a dimer that binds sequentially and cooperatively to the O_R1 and O_R2 sites of the $O_R O_{lac}$ promoter. Finally, both repression and activation, as implemented in the individual modules above, were placed together on one plasmid, to create a repressor–activator system (Fig. 1d).

A mathematical model was developed to account for the *in vivo* behaviour of this modular system, initially using equilibrium modelling of the underlying biochemical reaction scheme to capture the mean transcriptional response of the three regulated systems: the repressor-only, activator-only and repressor–activator systems. These considerations allow us to derive explicit expressions for the probabilities of each of the six binding states as a function of the inducer levels (see Supplementary Information). These expressions were globally fitted to the transcriptional induction data using the unregulated system as a baseline, leading to the mean fluorescence results (blue lines) shown in Fig. 1f–h.

The deterministic model was then extended to include stochastic effects²⁵. Preliminary investigations with the stochastic model revealed that fluctuations in the concentrations of transcription factors have very minor effects on the variability in the expression levels of the GFP reporter protein. Therefore, synthesis, degradation, and multimerization of CI and LacI were not included in our baseline model. All stochastic modelling was carried out using the BioNetS software²⁶, which uses a highly optimized version of the Gillespie Monte Carlo algorithm. To create the distributions of the GFP reporter protein, synthesis and degradation of GFP messenger RNA and protein were tracked explicitly. To include cell growth and division in the model, the cell volume was treated as a random variable that undergoes exponential growth²⁶ with a mean doubling time of 20 minutes, as observed in the experimental system. At cell division, the volume is halved and the mRNA and GFP molecules are divided between daughter cells on the basis of a binomial distribution²¹.

Because the plasmid used in our experimental system has a high-copy ColE1 origin of replication, intercellular copy-number variability can be large²⁷. Rather than model the complex copy-number control system explicitly, we assumed a gamma distribution at the time of cell division (see Supplementary Information). The mean was assumed to be 50 plasmid copies per cell²⁸. The variance was adjusted so that the model accurately captures the fluctuations in GFP levels observed in the activator-only system, and then kept at this value

¹Department of Biomedical Engineering, Bioinformatics Program, Center for BioDynamics and Center for Advanced Biotechnology, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA. ²Department of Statistics and Operations Research, ³Department of Mathematics, ⁴Department of Pharmacology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁵Institute for Optical Sciences and Department of Chemical and Physical Sciences, University of Toronto at Mississauga, 3359 Mississauga Rd, Mississauga, Ontario L5L 1C6 Canada. ⁶Department of Bioengineering, University of California, San Diego, 9500 Gillman Drive, La Jolla, California 92093, USA.

*These authors contributed equally to this work.

when the repressor-only and repressor-activator systems were considered. Comparison of the experimental results and the stochastic simulations in Fig. 1e and Fig. 1i–k shows that the model can accurately capture the *in vivo* behaviour of the unregulated system and all three regulated systems.

The coefficients of variation (CVs) of the three regulated systems are shown in the insets of Fig. 1f–h. As can be seen, the stochastic model slightly underestimates the experimentally observed variability. This discrepancy of approximately 0.1 in CV for each system is largely due to the model's inability to capture the long tails that extend into high fluorescence levels in the experimental flow cytometry histograms, and indicates that our relatively simple model does not account for all sources of variability. For example, while the model includes extrinsic noise⁵ that arises from cell growth and division and fluctuations in plasmid copy number, it does not account for cell-to-cell variability in the transcriptional and translational machinery.

An important test for any mathematical model is its ability to predict the outcome of novel experiments. We designed and conducted several new experiments on the basis of the model's

predictions. First, the stochastic model results revealed that a large amount of the variability in GFP levels is attributable to fluctuations in the plasmid copy number. To test this prediction, we placed the repressor-activator system on a low-copy plasmid that is tightly regulated, maintaining 3–4 copies of the plasmid per cell²⁸. Experiments were then conducted using both arabinose and IPTG as an inducer, and as expected, the cells with the low copy number showed reduced expression levels. In the stochastic model, the plasmid copy number was fixed at three, and the concentration of LacI was proportionally reduced to reflect this change. The functional relationship between arabinose level and CI concentration was taken to be the same in the low-copy case as in the high-copy case, to match our experimental observation that reducing the plasmid copy number did not affect the shape of the system's arabinose response curve (see Supplementary Information). No additional changes were made to the model. As seen in Fig. 2, the experimental distributions for GFP expression levels closely match those predicted by the model, indicating that the model accurately captures fluctuations in expression levels that are attributable to varying plasmid copy number.

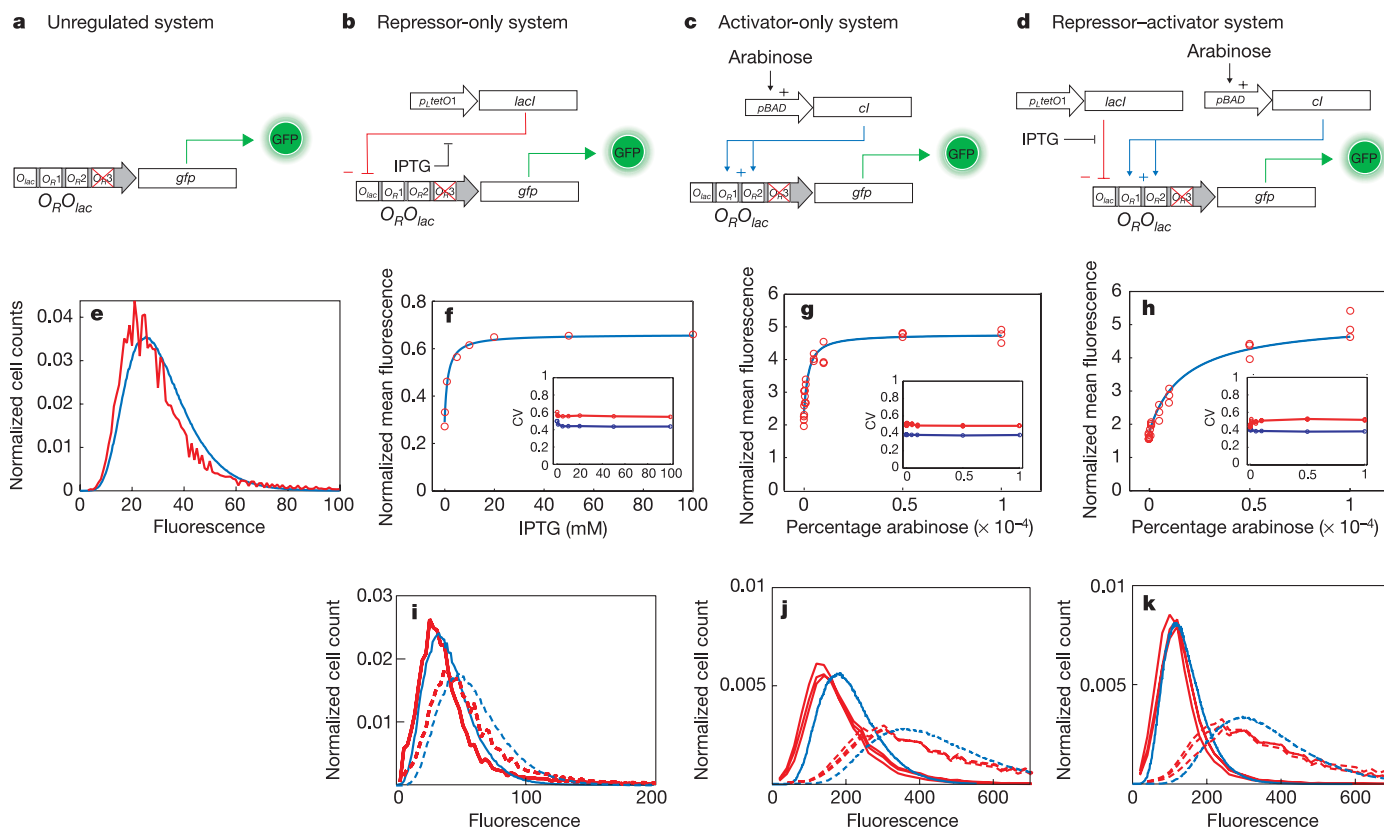


Figure 1 | Unregulated, repressor-only, activator-only and repressor-activator systems on high-copy plasmids. Schematic designs, experimental data and model data are shown. **a**, The unregulated system with the engineered $O_R O_{lac}$ promoter controlling the reporter *gfp*. The white boxes in the $O_R O_{lac}$ promoter represent the operator sites at which CI and LacI proteins bind. The red cross at O_R3 represents the point mutation that reduces CI binding at that operator site. **b**, The repressor-only system consists of the P_{tetO1} constitutive promoter controlling *lacI* and the $O_R O_{lac}$ promoter controlling *gfp*. **c**, The activator-only system consists of the $pBAD$ promoter controlling *ci* and the $O_R O_{lac}$ promoter controlling *gfp*. **d**, The repressor-activator system represents a combination of the repressor-only and activator-only systems. **e**, Histograms of the unregulated system: experimental data (red) and stochastic model data (blue). The x axis represents arbitrary fluorescence units from flow cytometry, and the y axis represents the frequency of cells producing the corresponding fluorescence level. **f**, Repressor-only system results: GFP expression represented as

normalized fluorescence versus IPTG level; red circles are experimental data and the blue lines are the results of the deterministic model. The inset shows CV versus IPTG level: experimental data (red) and stochastic model data (blue). **g**, Activator-only system results: normalized fluorescence versus arabinose level, with an inset showing CV versus arabinose level. **h**, Repressor-activator system results: normalized fluorescence versus arabinose level. The inset shows CV versus arabinose level, with 10 mM IPTG in each case. **i**, Histograms of normalized cell counts versus arbitrary fluorescence units, of experimental data (red) and stochastic model data (blue) for the repressor-only system. The solid lines in the histograms are the results for no inducer (IPTG in this case) and the dashed lines are the results for the highest level of IPTG. **j**, Histograms (as above), showing results for no inducer (solid lines) and the highest level of arabinose (dashed lines) for the activator-only system. **k**, Histograms (as above), showing results for no arabinose (solid lines) and the highest level of arabinose (dashed lines), with 10 mM IPTG in each case for the repressor-activator system.

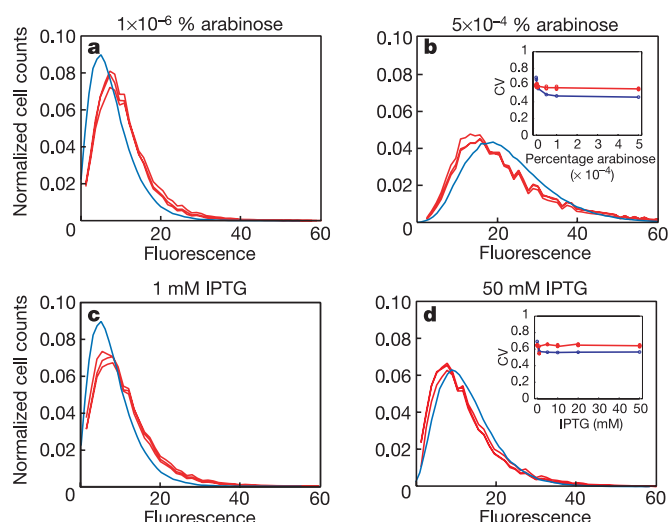


Figure 2 | Histograms of model data (blue lines) and experimental data (red lines) for the repressor-activator system on a low-copy plasmid. The x axis represents arbitrary fluorescence units from flow cytometry, and the y axis represents the frequency of cells producing the corresponding fluorescence level for: **a**, 1×10^{-6} % arabinose, no IPTG; **b**, 5×10^{-4} % arabinose, no IPTG; **c**, 1 mM IPTG, no arabinose; **d**, 50 mM IPTG, no arabinose. The inset in **b** shows CV versus arabinose level, and the inset in **d** shows CV versus IPTG level; model data (blue lines) and experimental data (red lines).

The insets of Fig. 2b and d show a comparison of the CVs for the experimental and model results. The model accurately predicts the behaviour of the CV for the low-copy case, taking into account the 0.1 discrepancy discussed above. At very low inducer levels, the model predicts a larger CV than is observed experimentally. We attribute this discrepancy to autofluorescence in the experimental system. This has the effect of artificially increasing the

mean fluorescence, thereby decreasing the CV. The effect is only important at low fluorescence levels, where autofluorescence makes up a significant portion of the measured mean value.

To test the model's predictive power in a more complex system, we added positive feedback to the repressor-activator system. To accomplish this, we expanded the system so that the *cl* gene was transcribed polycistronically with *gfp* under the control of the $O_R O_{lac}$ promoter (Fig. 3a). Adding feedback to the stochastic model requires the explicit inclusion of the CI mRNA and protein molecular abundances and the biochemical processes that affect their levels (synthesis, degradation, multimerization). The parameters that control these processes were chosen so that in the absence of feedback the expanded model produced CI protein levels identical to that of the simpler model (see Supplementary Information). As can be seen in Fig. 3b–g, the agreement between the behaviour predicted by the model and the experimental results is very good, validating the bottom-up approach to understanding gene regulatory networks.

We also used the model to conduct a systematic analysis of how the different sources of noise contribute to the overall variability (Supplementary Fig. 7). Surprisingly, we found that if we run simulations with no cell growth or division using an ensemble of cells with varying plasmid copy number, the CV increases above the level at which cell growth and division proceed normally. This result was unexpected, because cell growth and division are generally thought to add variability to expression levels. A theoretical explanation of this counterintuitive prediction is provided in the Supplementary Information. In very general terms, this phenomenon can be understood as follows: in a population of cells with varying plasmid copy number, the intercellular variability increases as expression levels increase from new protein synthesis. Cell division not only limits the mean protein level, it also causes the distribution to be more tightly centred about the mean (Supplementary Fig. 8). This leads to an overall reduction of the variability in protein expression for the high-copy system where plasmid copy-number fluctuations account for a significant portion of the noise. In

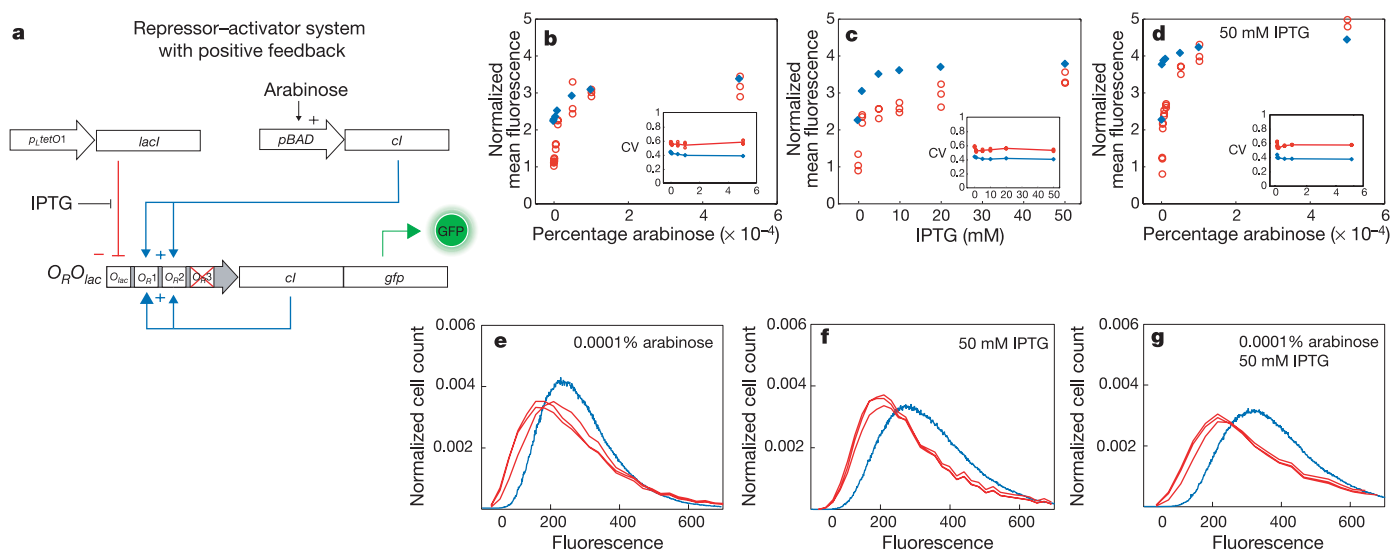


Figure 3 | Repressor-activator system with positive feedback.

a, Schematic of the positive feedback construct, where the *cl* gene is added to the repressor-activator system on a high-copy plasmid. The *cl* gene is incorporated, along with *gfp*, in a polycistronic region controlled by the engineered $O_R O_{lac}$ promoter. Model predictions and experimental results for the repressor-activator system with positive feedback: **b**, Normalized mean (arbitrary fluorescence units) versus arabinose level, where the red circles are experimental data and the blue diamonds are model predictions. The inset shows CV versus arabinose level. **c**, Normalized mean versus IPTG

level, with an inset showing CV versus IPTG level. **d**, Normalized mean versus arabinose level, with 50 mM IPTG in each case, with an inset showing CV versus arabinose level for 50 mM IPTG. **e**, Histograms, normalized cell counts versus arbitrary fluorescence units, of experimental data (red) and model predictions (blue) for 0.0001% arabinose. **f**, Histograms of experimental data (blue) and model predictions (red) for 50 mM IPTG. **g**, Histograms of experimental data (blue) and model predictions (red) for 0.0001% arabinose and 50 mM IPTG.

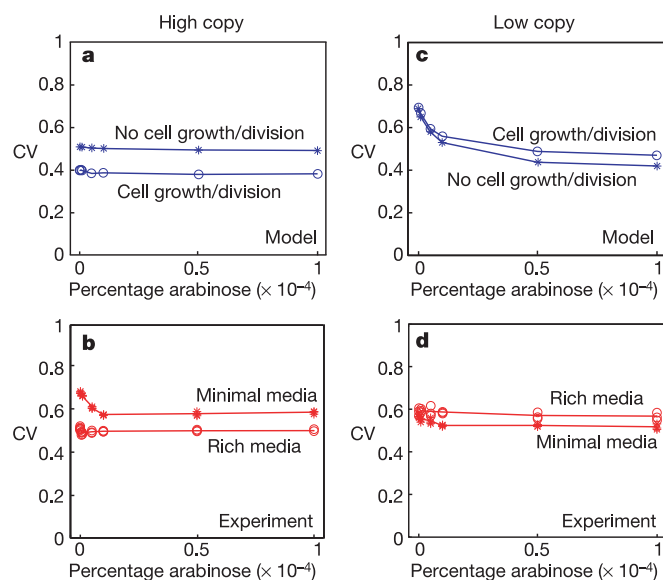


Figure 4 | CV versus arabinose levels for the repressor-activator system. **a**, Model results for the high-copy case with cell growth and division (circles) and with no cell growth or division (stars). **b**, Experimental results for the high-copy case for cells in rich media (circles) and minimal media (stars). **c**, Modelling results for the low-copy case with cell growth and division (circles) and with no cell growth or division (stars). **d**, Experimental results for the low copy case for cells in rich media (circles) and minimal media (stars).

contrast, for the low-copy case where plasmid copy number is tightly controlled, the model predicts that cell growth and division should not reduce noise levels and instead, halting cell growth and division is expected to decrease the variability.

We considered ways to experimentally test these model predictions. Most methods of achieving arrest of cell growth and division (such as addition of the drug cephalixin²⁹) cause drastic changes in cell morphology or protein production, both of which would be detrimental to our experimental design. We found cell behaviour in minimal media to be the best approximation of our model conditions. In minimal media, cell growth and division slow considerably, while protein overproduction from plasmids can continue.

Cells were placed in M9 minimal media and 2% glucose with incubator conditions identical to those in the rich media experiments. Our model predicts that the variability in protein expression levels can increase if cells with the repressor-activator system on a high-copy plasmid do not grow or divide (Fig. 4a). The results in Fig. 4b show that the cells in minimal media with high-copy plasmids do in fact have a higher CV than those that are growing and dividing regularly in rich *Luria* broth (LB) media. Additionally, our model predicts that the variability in expression levels for the low-copy case decreases when cell growth and division are stopped (Fig. 4c). This effect is validated experimentally in Fig. 4d where the results show that cells with low-copy plasmids in minimal media have lower CVs than similar cells growing and dividing in rich media. (See Supplementary Information for similar results with stationary phase cells.)

Previous work in synthetic biology^{1–23} has shown that one can develop a mathematical model for a given synthetic gene network and use the model to describe the behaviour of that network. These efforts set the stage for the possibility of using quantitative models of regulatory subsystems to predict the behaviour of larger, more complex regulatory networks, one of the main goals of synthetic biology. In this study, we show that such a bottom-up approach is indeed achievable and can lead to insights and testable predictions that are biologically meaningful. Specifically, we show that we can fit

a stochastic model to experimental data from a modular, synthetic gene network, and then extend the model and use it to predict quantitatively the behaviour of a more complex network involving regulatory feedback. The model also accurately predicted the effects of changing plasmid copy numbers on the experimentally observed levels of variability in GFP expression and suggested that halting cell growth and division in cells with high-copy plasmids should increase fluctuations in protein expression, which was confirmed experimentally. This synthesis of theory and experiment has much to offer in investigations of cellular behaviour, and such approaches may eventually allow us to assemble a genuine bottom-up picture of the intricate processes of gene regulation.

METHODS

Plasmid construction. Plasmid construction began with the pZE21-MCS1 modular construct of Lutz and Bujard²⁸. This plasmid backbone contains the *P_{tet}O1* promoter, *T1T2* terminator, kanamycin resistance gene, *ColE1* origin of replication and multiple cloning sites. The low copy origin of replication was from the pZS*24-MCS1 plasmid²⁸. To create the *O_RO_{lac}* promoter, the *O_R* region of bacteriophage λ was altered with the G-to-Tor3-r2 point mutation in the *O_R3* operator site²⁴ (greatly reducing its affinity for the CI protein), and an *O_{lac}* operator site was added upstream of the *O_R* region. The promoter sequence is shown in Supplementary Fig. 12. The point mutation and operator insertion were carried out by polymerase chain reaction (PCR; using PfuTurbo DNA polymerase from Stratagene and an MJ Research PTC-100 thermal cycler) with primers designed such that the point mutation and operator site were placed in the primers.

The *lacI* gene was PCR-amplified, with PfuTurbo and an MJ Research thermal cycler, from pTrc99a, the *clts* and *O_R* region were from pGW7 (ATCC), the *gfpmut3* was from pJBA111 (J. B. Andersen, Technical University of Denmark), and the pBAD promoter was from pBADHisA (Clontech).

Cell growth and expression experiments. All plasmids were inserted into the *E. coli* strain JM2.300 ($-\lambda$, *lacI22 rpsL135* (StrR), *thi-1*) via the transformation and storage solution heat-shock transformation protocol³⁰. Cells for experimentation were grown overnight and cultures were inoculated 1:300 in LB media and with 0.03 mg ml⁻¹ kanamycin antibiotic. Varying levels of IPTG and arabinose were added separately and in combination to affect the behaviour of the inducible promoters. Cultures were grown in a 37°C incubator shaking at 300 r.p.m. for approximately three hours, at which point the cells reached a steady state (see Supplementary Information), and the absorbance at 600 nm (*A_{600 nm}*) was between 0.3 and 0.4. The minimal media cultures used the same conditions as LB cultures except that the minimal media cultures were grown in M9 minimal media with 2% glucose.

Data acquisition and analysis. Culture samples were spun down at 8,000 r.p.m. to pellet cells, and supernatant was removed. The samples were then resuspended in 0.75 ml of 1 × PBS, and fluorescence measurements were taken via flow cytometry. The Becton Dickinson FACScalibur flow cytometer was used to measure 50,000 cells of each sample from a culture representing one level of inducer. Excitation of GFP was achieved via a 488-nm argon excitation laser and fluorescence measured with the 515–545 nm emission filter. The flow cytometer generates logscale values using a 10-bit analogue-to-digital converter, yielding integers in the range 0 to 1,023 for each of three measurements: fluorescence intensity, forward-scattering, and side-scattering. Cells were collected within a small forward-scatter and side-scatter gate, in the Cellquest software designed by BD Biosciences, to minimize fluorescence variation due to cell size.

Modelling and stochastic simulations. The model parameters were estimated using Matlab's least-squares fitting routine. The mean expression levels were fitted to the repressor-only, activator-only and repressor-activator system results. To initialize the least-squares routine, we used 2,000 random initial guesses for the model parameters, generated from a uniform distribution within a biologically reasonable regime of parameter space. A discussion of the sensitivity of the model's output to the values of the parameters is given in the Supplementary Information. The best parameter set was chosen from those 2,000 runs. All the fits were based on data normalized to the mean GFP expression level of the unregulated system. All stochastic simulations were done in BioNetS²⁶ using the Gillespie algorithm. See Supplementary Information for details.

Received 7 October; accepted 18 November 2005.

1. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

2. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
3. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
4. Becskei, A., S  raphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
5. Elowitz, M., Levine, A., Siggia, E. & Swain, P. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
6. Ozbudak, E., Thattai, M., Kurtser, I., Grossman, A. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
7. Rosenfeld, N. Y., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**, 785–793 (2002).
8. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1407–1470 (2002).
9. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA* **100**, 7714–7719 (2003).
10. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
11. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behaviour in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
12. Weiss, R. et al. Genetic circuit building blocks for cellular computation, communications, and signal processing. *Natural Comput.* **2**, 47–84 (2003).
13. Basu, S., Mahreja, R., Thiberge, S., Chen, M. T. & Weiss, R. Spatiotemporal control of gene expression with pulse generating networks. *Proc. Natl. Acad. Sci.* **101**, 6355–6360 (2004).
14. You, L., Cox, R. S. III, Weiss, R. & Arnold, F. H. Programmable population control cell-cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
15. Kramer, B. P. et al. An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnol.* **22**, 867–870 (2004).
16. Kobayashi, H. et al. Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci.* **101**, 8414–8419 (2004).
17. Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnol.* **22**, 841–847 (2004).
18. Fung, E. et al. A synthetic gene-metabolic oscillator. *Nature* **435**, 118–122 (2005).
19. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. USA* **102**, 3581–3586 (2005).
20. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
21. Rosenfeld, N. Y., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single cell level. *Science* **307**, 1962–1965 (2005).
22. Isalan, M., Lemerle, C. & Serrano, L. Engineering gene networks to emulate *Drosophila* embryonic pattern formation. *PLoS Biol.* **3**(3), e64 (2005).
23. Kramer, B. P. & Fussenegger, M. Hysteresis in a synthetic mammalian gene network. *Proc. Natl. Acad. Sci. USA* **102**, 9517–9522 (2005).
24. Ptashne, M. *A Genetic Switch: Phage λ and Higher Organisms* (Cell Press & Blackwell Scientific, Cambridge, Massachusetts, 1992).
25. Kepler, T. & Elston, T. C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
26. Adalsteinsson, D., McMillen, D. & Elston, T. C. Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. *BMC Bioinform.* **5**, 24 (2004).
27. Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: plasmid copy number control. *Q. Rev. Biophys.* **34**, 1–59 (2001).
28. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-12 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
29. Greenwood, D. & O'Grady, F. Comparison of the response of *Escherichia coli* and *Proteus mirabilis* to seven β -lactam antibiotics. *J. Infect. Dis.* **128**, 211–222 (1973).
30. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, New York, 1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NIH, NSF and DARPA.

Author Contributions T.C.E. and J.J.C. are co-senior authors.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.J.C. (jcollins@bu.edu).

Origins of extrinsic variability in eukaryotic gene expression

Dmitri Volfson^{1,2*}, Jennifer Marciniak^{1*}, William J. Blake³, Natalie Ostroff¹, Lev S. Tsimring² & Jeff Hasty¹

Variable gene expression within a clonal population of cells has been implicated in a number of important processes including mutation and evolution^{1,2}, determination of cell fates^{3,4} and the development of genetic disease^{5,6}. Recent studies have demonstrated that a significant component of expression variability arises from extrinsic factors thought to influence multiple genes simultaneously^{7–10}, yet the biological origins of this extrinsic variability have received little attention. Here we combine computational modelling^{11–18} with fluorescence data generated from multiple promoter–gene inserts in *Saccharomyces cerevisiae* to identify two major sources of extrinsic variability. One unavoidable source arising from the coupling of gene expression with population dynamics leads to a ubiquitous lower limit for expression variability. A second source, which is modelled as originating from a common upstream transcription factor, exemplifies how regulatory networks can convert noise in upstream regulator expression into extrinsic noise at the output of a target gene⁹. Our results highlight the importance of the interplay of gene regulatory networks with population heterogeneity for understanding the origins of cellular diversity.

To investigate variability in eukaryotic gene expression, we used the native *GAL1* promoter of the yeast *Saccharomyces cerevisiae* with *yEGFP* (yeast-enhanced green fluorescent protein) as a quantifiable marker. We constructed five yeast strains by varying the number of copies integrated into the *GAL1-10* locus on chromosome II. These

multiple-copy constructs can be used to determine whether the variability in gene expression is due to intrinsic or extrinsic sources. Adopting the standard nomenclature, intrinsic noise originates from the small number of regulatory molecules participating in the inherently noisy biochemical reactions leading to expression, whereas extrinsic variability arises from sources such as the variation of the elements of the transcriptional machinery common to all genes or fluctuating environmental variables. This extrinsic variability can have stochastic or deterministic origins, and we reserve the use of the term noise to denote a random, as opposed to a deterministic, origin for the variations.

In the absence of glucose, the *GAL1* promoter is activated in response to galactose, and our cell strains were induced to produce *yEGFP* using varying amounts of galactose ranging from 0.1% to 2%. Single-cell fluorescence data were collected using flow cytometry, and representative distributions are depicted in Fig. 1a for fixed galactose concentration. Information regarding the underlying transcriptional process can be extracted by analysing the scaling properties of the data sets with respect to the copy number. For example, the mean fluorescence increases as a function of galactose concentration and the copy number (Fig. 1c), and when fluorescence is divided by the corresponding copy number, the induction curves for all of the strains ‘collapse’ to a single curve (Fig. 1c, d). This indicates that each promoter–gene pair transcribes at the same mean rate upon insertion.

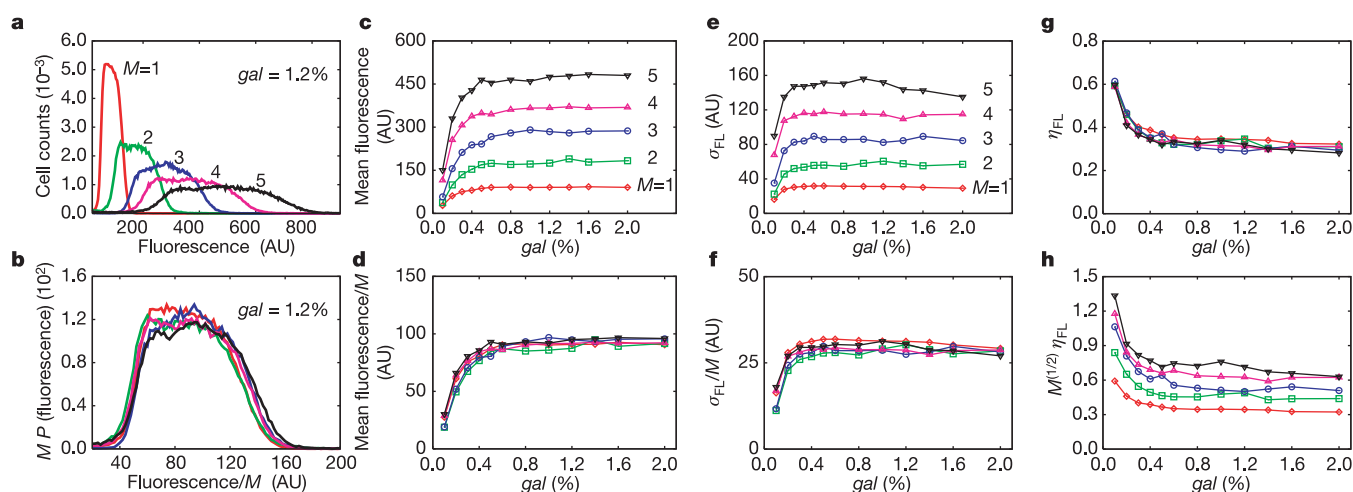


Figure 1 | Experimental results for GFP expression for different copy numbers and galactose concentrations. **a**, Histogram of GFP measurements for copy numbers from $M = 1$ to $M = 5$ above the saturation ($gal = 1.2\%$). **b**, The collapse of GFP distributions under the transformation $F \rightarrow F/M$, $P(F) \rightarrow MP(F/M)$ implies an extrinsic source of variability. AU, arbitrary units. **c**, Induction curves for copy numbers from $M = 1$ to $M = 5$.

d, Collapse of the induction curves implies that transcription from each promoter is independent. **e**, Standard deviations of GFP corresponding to induction curves. **f**, The collapse of the standard deviation implies an extrinsic source of variability. **g**, The collapse of the coefficient of variation for different copy number implies an extrinsic source of variability. **h**, Lack of collapse implies that the variability is not of intrinsic origin.

¹Department of Bioengineering, and ²Institute for Nonlinear Science, University of California San Diego, La Jolla, California 92093, USA. ³Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA.

*These authors contributed equally to this work.

We can use scaling in an analogous manner to deduce information about the origins of variability in the system. Depending on the relative magnitude of intrinsic noise versus extrinsic variability, we expect to observe different dependencies (scalings) on the copy number M .

In order to illustrate how scaling can be used to determine whether the observed variability is intrinsic or extrinsic, suppose there are M promoter–gene pairs producing GFP molecules, and let the level of production of the i th copy be the sum of two terms representing the mean and fluctuations about the mean: $g_i = \langle g_i \rangle + \tilde{g}_i$. The time-average of the fluctuating term is zero ($\langle \tilde{g}_i \rangle = 0$) and the variance is given by $v_i = \langle g_i^2 \rangle - \langle g_i \rangle^2$. Then the mean and variance for the total amount of GFP produced by M copies are $\langle G_M \rangle = \sum_{i=1}^M \langle g_i \rangle$ and $V_M = \sum_{i=1}^M \left[v_i + \sum_{j \neq i} \langle \tilde{g}_i \tilde{g}_j \rangle \right]$, respectively. The first term in the expression for V_M represents the sum of the variances of the individual copies, and the second term is the contribution of cross-correlations between them. For identical copies, $\langle g_i \rangle = \langle g \rangle$, $v_i = v$, $\langle \tilde{g}_i \tilde{g}_j \rangle = c$ and the total mean is proportional to the copy number, $\langle G_M \rangle = M \langle g \rangle$, whereas the total variance is given by $V_M = Mv + M(M-1)c$. When $c = 0$, the fluctuations are completely uncorrelated and the variance is proportional to the copy number. This is the intrinsic noise limit, where expression events from individual copies are completely independent. When $c = v$, the variance is proportional to the square of the copy number. In this limit the fluctuations are completely correlated and often caused by extrinsic factors. Using the results for the variance and mean, the coefficient of variation η (standard deviation divided by the mean) is expected to scale as $\eta \propto M^{-1/2}$ for the case of purely intrinsic noise and be independent of the copy number in the case of extrinsic variability. Thus, if the system is dominated by intrinsic factors, the quantity $M^{1/2}\eta$ should collapse to the same galactose-dependent curve for all copy numbers. On the other hand, if the system is dominated by extrinsic sources, then η itself should collapse for all copy numbers.

In order to ascertain whether the GAL system is dominated by intrinsic noise or extrinsic variability, we plot the unscaled (η) and scaled ($M^{1/2}\eta$) coefficient of variation as a function of galactose (Fig. 1g, h). We observe that even though different strains exhibit small differences, the collapse of the data in the unscaled case indicates that the coefficient of variation is independent of the copy number. An even more striking collapse is presented in Fig. 1b, where we rescale fluorescence distributions $P(F)$ for a representative value of galactose concentration above saturation ($gal = 1.2\%$) and for different strains. In obtaining this collapse, we first normalized the distributions and rescaled by the copy number as follows: $F \rightarrow F/M$, $P(F) \rightarrow MP(F/M)$. This collapse implies that not only is the standard deviation proportional to the copy number, but generally the n th moment of the fluorescence distribution is proportional to M^n . These results corroborate findings by a recent study⁸ that used a ‘two-colour’ approach⁷ to demonstrate that extrinsic sources dominate the variations observed in the GAL system. In the Supplementary Information, we demonstrate that our findings are unchanged when the genes are moved to different chromosomes, and we describe how our data analysis can be generalized for intermediate cases where there is mixture of scaling arising from the contribution of both intrinsic and extrinsic sources.

Our analysis of the experimental data provides an important constraint when attempting to deduce the biological origin of the variability, because a non-trivial test of any model description is that it must exhibit correct copy-number scaling. We first construct a fully deterministic model that couples a mesoscopic description of cell growth and division with a microscopic description of gene expression (Box 1, model A). For colony growth, we generalize the classical model introduced by ref. 19, which is known to capture the major experimental observations regarding the cell size distributions. We assume that each cell grows exponentially at a rate α and divides after reaching the same size x_0 . The division is asymmetric so that the size of a mother cell after division is ξ/κ larger than that of a daughter

cell ($\xi + \kappa = 1$, $\xi > \kappa$). After a transient, such a colony reaches a stationary distribution of sizes with two characteristic peaks corresponding to subpopulations of daughter and mother cells (see Fig. 2a, where the theoretical distribution is compared with experimental data obtained from the flow cytometry forward scatter data). In addition to the process of growth and division, model A assumes that each cell is producing GFP at a constant rate γ , and GFP is distributed at division in the same proportion as the volume. We show in the Supplementary Information that the joint probability of finding a cell of a given size x , and with GFP content g , has the functional form $\mathcal{P}_{vg}(x, g) = \mathcal{P}_v(x) \mathcal{P}_g(g/a, \kappa)$, where $a = \gamma/\alpha$ is the ratio of the GFP production rate and volume growth rate. Using this result, we then derive an explicit expression for the distribution $\mathcal{P}_g(g/a, \kappa)$ (see Fig. 2b), and from there generate expressions for the mean and standard deviation of the stationary distribution of GFP, $\langle g \rangle = a$ and $\sigma_g = a\eta_g(\kappa)$, where $\eta_g(\kappa) = [1 - \frac{2}{3\kappa\xi}(\kappa^2 \ln(\kappa)^2 + \kappa\xi \ln(\kappa) \ln(\xi) + \xi^2 \ln(\xi)^2)]^{1/2}$.

The model results for the mean and standard deviation have non-trivial implications consistent with our experimental findings. Because the rate of GFP production $\gamma = \alpha a$ is proportional to the copy number M , the functional form $\mathcal{P}_g(g/a, \kappa)$ implies that distributions of GFP for different copy numbers M should collapse after the transformation $g \rightarrow g/M$, $\mathcal{P}_g \rightarrow M\mathcal{P}_g$. In particular, because the mean value of the GFP distribution is proportional to the copy number and the standard deviation is proportional to the mean, the coefficient of variation is independent of M as observed in the experiments. In addition, the model yields an explicit dependence of the coefficient of variation on the partition ratio κ (see Supplementary Information), and using a previously reported value for this parameter²⁰ we obtain a coefficient of variation that is in close quantitative agreement with our experiments for high concentrations of galactose (Fig. 2e, straight line). Importantly, this

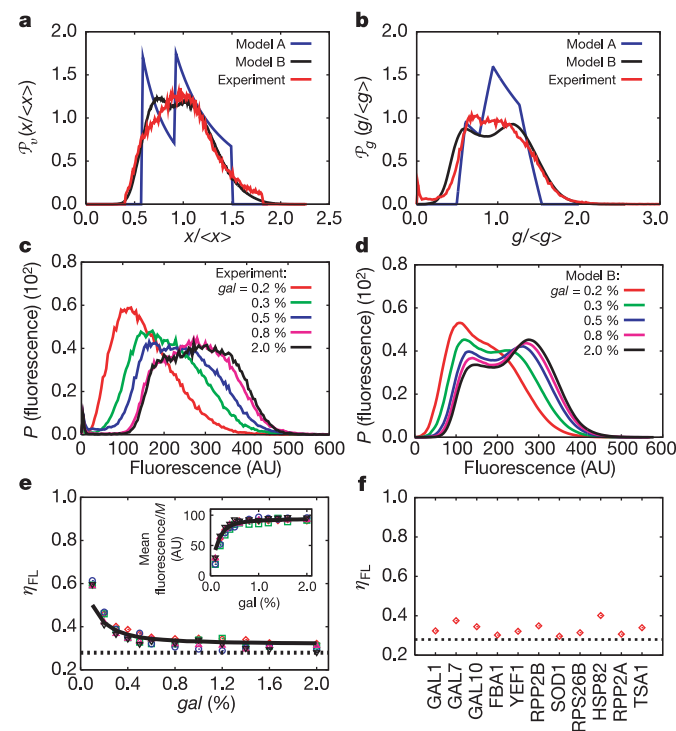


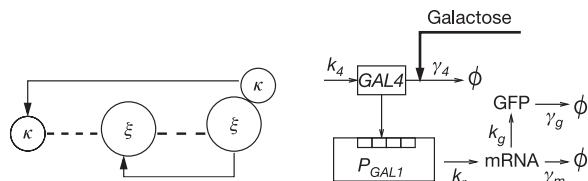
Figure 2 | Comparison of models A and B with experimental findings. **a**, Distributions of cell sizes. **b**, Distributions of GFP at fixed galactose concentration and $M = 1$. **c**, Distributions of GFP for $M = 3$ and for a number of galactose concentrations. **d**, The same as **c** but from model B. **e**, Collapse of the induction curves (symbols) along with the results of simulations of model B (black solid line). **f**, Validation of model predictions. The model noise floor (dashed black line) along with experiments using different fusions (symbols) are shown.

Box 1 | Coupling gene expression to population dynamics

Our two models are constructed by coupling a subsystem for gene expression from the *GAL1* promoter to a subsystem representing the growth and division of individual cells in the population. For the gene expression subsystem, we introduce a simple scheme analogous to the previously reported model¹⁵. In our model each of the M identical promoters may be in either an inactive state (O_1) or an active one (O_2). The latter represents a preinitiation complex with all the components necessary for transcription in place. The production of mRNA proceeds from state O_2 with rate k_r . After mRNA has been produced, translation ensues with rate k_g . Both mRNA and GFP are allowed to decay at rates γ_m and γ_g , respectively. The effect of the inducer is modelled by taking into account the presence of Gal4p proteins, which are responsible for the activation of the P_{GAL1} promoter. These proteins dimerize and bind to four binding sites of P_{GAL1} , activating the process of transcription. When multiple copies of the promoter–gene pair are inserted, they share the pools of Gal4p, mRNA and GFP. We assume that Gal4p monomers are constitutively produced with rate α_4 . Another protein (Gal80p) binds to Gal4p and renders it unable to activate the promoter. We model this by taking the degradation rate of Gal4p (γ_4) to be inversely proportional to the galactose concentration. Combining these assumptions, the differential equations describing the deterministic model for gene expression may be deduced from the chemical reactions. The rate equations for Gal4p, mRNA and GFP, respectively, read.

$$\dot{g}_4 = \alpha_4 - \gamma_4 g_4, \quad \dot{r} = Mk_r H(g_4) - \gamma_m r, \quad \dot{g} = k_g r - \gamma_g g$$

where $H(g_4) = Kg_4^z / (1 + Kg_4^z)$ is the Hill function for the activation process, and we use the cooperativity $z = 8$, which stems from the binding of four Gal4p dimers; K is a lumped parameter that describes the net effect of dimerization/dissociation and binding/unbinding; and the dependence on the galactose concentration arises from $\gamma_4 \propto gal^{-1}$ (see Supplementary Information for a description of our modelling methods, along with a derivation of this model from a more general model of the galactose signalling pathway).



For model A, our primary motivation is to elicit the extrinsic variability arising from cell growth and division, so we first develop a deterministic model for the gene expression component. The characteristic timescale here is set by the growth rate of the cells. The stability of GFP implies a negligible degradation rate γ_g , and because the rates of degradation of Gal4p and mRNA are fast compared with the cellular growth rate, we can eliminate the dynamical equations for g_4 and m . Using these considerations, the

dynamics of GFP are then reduced to production at a constant rate, $\dot{g} = \Gamma = M[k_g k_r H(\alpha_4/\gamma_4)/\gamma_m]$.

We next couple a deterministic model for cellular growth and division with the deterministic production of GFP. To account for population dynamics we consider an asynchronous population of cells, each growing exponentially with rate α and dividing in an asymmetric fashion¹⁹ after reaching a certain maximum size. We let the size of a mother cell after division be ξ/κ times larger than that of the daughter cell, where $\kappa < 1/2(\xi + \kappa = 1)$. The process of growth and division repeats indefinitely, producing a growing structured population of cells of multiple generations where each cell is described by a position within the cell cycle and genealogical age. The coupling between this growth model and gene expression arises through the partitioning of GFP content at division. The system of evolution equations describing the size (x) and number of GFP molecules (g) is given by

$$\dot{x} = \alpha x, \quad \dot{g} = \Gamma$$

which should be supplemented by the renewal equations (division rules), according to $x_0 \rightarrow \kappa x_0 + \xi x_0, g(x_0) \rightarrow \kappa g(x_0) + \xi g(x_0)$ where $x_0, g(x_0)$ are the cell size and GFP content at the moment of division, respectively. The additional assumptions behind this model may be summarized as follows. We distinguish only two generations of cells—daughters and mothers—thus neglecting the difference between mothers of second generation and higher. We assume that cells are growing with the same rate regardless of their age, and that both daughters and mothers divide after reaching the same size and in the same proportion. Finally, we assume that GFP is distributed among daughter cells in the same proportion as the size (fast diffusion of GFP molecules). Overall, the statistics of the population is modelled in terms of the equation for the joint probability $\mathcal{P}_{vg}(x, g)$ to find a cell of a given size x and with a certain amount of GFP g , which is solved analytically (see Supplementary Information).

Model B generalizes model A by including several realistic features of both gene expression and population dynamics. For gene expression we use a hybrid approach whereby production and degradation of Gal4p proteins is modelled stochastically using the Gillespie algorithm²⁴, whereas production of mRNA and GFP is simulated with the set of ordinary differential equations introduced above. Consistent with our experimental observations, our hybrid approach eliminates all intrinsic fluctuations except for those induced by Gal4p monomers. This intrinsic noise is then transferred to the level of mRNA production through fluctuations in the production rate of mRNA, which is proportional to the Hill function $H(g_4)$. Here, the intrinsic noise manifests itself as extrinsic fluctuations common to all M copies. For the population dynamics we introduce genealogical age by using different values of κ depending on the generation²⁰, and introduce a stochastic component in division by drawing a cell size at division from a narrow gaussian distribution centred at x_0 . In addition, upon division the amounts of Gal4p, mRNA and GFP are distributed among the offspring in a binomial fashion.

result was obtained without any additional fitting (in this model, the coefficient of variation is shown to be a function of κ only). As an additional confirmation of the role of the population dynamics in genetic variability we calculated the coefficient of variation in gated subpopulations of cells within narrow windows of sizes and found that gating reduces the coefficient of variation by as much as 50% (see Supplementary Information).

Because the variability in the model arises from the coupling of population effects and deterministic gene expression, an interesting model prediction is that there is a 'noise floor' that provides the lower limit for expression variability. If there are no other sources of variability, this implies that for various highly expressed genes the observed variability should be similar. We tested this prediction by constructing GFP fusions to ten highly expressed *S. cerevisiae* genes. As predicted, we observed a near-constant level of variability for all of the constructed fusions (Fig. 2f). This level is consistent with the lower noise floor observed in the GAL system (Fig. 2e), and can be

taken as the lower limit for the variability in a growing population of cells. These results suggest that for high expression levels, population effects dominate and gene expression within a single cell is mostly deterministic.

Although our fully deterministic model A accounts for the majority of the observed variations in the inducible GAL1–GFP system, there are still certain limitations of this simple model. In particular, the theoretical GFP distribution shows a large peak that is not present in the experimental data (Fig. 2b). It occurs because model A ignores the stochastic fluctuations in the cell size at division, as well as the binomial nature of division of GFP among mother and daughter cells, and thus the GFP distribution is narrow. Furthermore, model A does not account for the increase of variability at small galactose values. We next generalize the model to describe the variability at small galactose values (model B; see Box 1). In this regime, the variability increases as the number of GFP molecules decreases. Although this dependence is consistent with an intrinsic noise source,

the scaling of the coefficient of variation with copy number implies an extrinsic source, and thus fluctuations in expression from uncorrelated *GAL1* promoters cannot be the origin. We therefore look upstream of the *GAL1* promoter. The *GAL1* promoter is activated by Gal4p protein dimers, so a common stochastic component of gene expression may originate from fluctuations in the level of the Gal4p activators. We tested the idea of activator-mediated fluctuations with a quantitative model incorporating the galactose-dependent degradation of Gal4p (Fig. 2a–c). The generalized model shows excellent agreement for the distributions and scaling of the coefficient of variation as a function of galactose. In addition, model B predicts that other Gal4p-regulated promoters should exhibit extrinsic scaling at low galactose, and we present experimental validation of this prediction in the Supplementary Information.

We have shown how experimental classification of variability can be used to constrain the space of possible models and lead to the development of a model incorporating variations arising from the coupling of population growth and gene expression. Although this dominant source of variability is likely to exist in all systems, the generalization of our model to include upstream activators of the *GAL1* promoter is system-specific. It would be interesting to explore other inducible eukaryotic systems to see whether the scaling of the variability also implies an extrinsic source for low induction levels. In the context of the reported correlations observed in the transcripts of proximally located genes²¹, our results suggest that the biological origin of these correlations could be a common upstream transcriptional regulator.

METHODS

Plasmids, yeast strains and growth conditions. All plasmid backbones were derived from pRS403, pRS404, pRS405 or pRS406 shuttle vectors (Stratagene). The *GAL1-10* promoter region and the *yEGFP* were obtained from pESC1-yG. All yeast strains (see Supplementary Table II) in this study are derived from the *S. cerevisiae* parent strain YPH500 (α , *ura3-52*, *lys2-801*, *ade2-101*, *trp1Δ63*, *his3Δ200*, *leu2Δ1*) (Stratagene). The strains were created by targeted chromosomal integration of shuttle vector constructs at either the *GAL1-10* locus on chromosome II or the *TRP1* locus on chromosome IV. All fusion strains were constructed as described previously²² (see also Supplementary Information). The yeast strains were grown in synthetic drop-out (SD) medium supplemented for selection of correct integrands containing 2% glucose at 30 °C. The strains were selected for the proper number of promoter–gene pairs by growing the cells on 2% galactose and assaying the cells for average fluorescence. All cloning steps were performed in *Escherichia coli* XL10-Gold (Stratagene).

Gene expression experiments. Exponentially growing yeast cells were diluted from 1:40 to 1:70 into medium containing 2% raffinose and a range of galactose concentrations (0.1–2.0% galactose) as inducer. After 14–18 h, the cells were assayed at mid-log phase at an absorbance at 600 nm of 0.55 ± 0.15 . Expression data were collected using a Becton-Dickinson FACSCalibur flow cytometer with a 488-nm argon excitation laser and a 515- to 545-nm emission filter (FL1) at a low flow rate. Forward scatter values and fluorescence values were collected for 100,000 cells. The standard list-mode files obtained from the flow cytometer were converted to ASCII format with MFI (E. Martz, <http://www.umass.edu/microbio/mfi>) and analysed using Matlab (The MathWorks, Inc.).

Modelling studies. Model A was solved analytically; details of the calculations are presented in the Supplementary Information. For the comparison with the experiment we used an effective ratio of the volumes of daughter cells after division to mother cells before the division, $\kappa = 0.34$, representing the weighted average of κ among several generations²⁰ (Fig. 1g). Model B was solved numerically using a hybrid technique²³ whereby the stochastic dynamics of Gal4p protein have been modelled with the Direct Gillespie algorithm²⁴, the intermediate reactions of dimerization and binding of Gal4p dimers were eliminated using quasi-stationary assumptions for the sake of simplicity, and the dynamics of mRNA and GFP were simulated using the rate equations presented in Box 1. For the population dynamics we used previously reported measurements regarding the structure of the population of *S. cerevisiae*²⁰. Namely, for the first generation, we used $\kappa = 0.4$ and for subsequent generations $\kappa = 0.3$; the mean size of the cells at divisions was chosen to be 1.089, 1.179 and 1.268 for genealogical ages from 2 to 4, respectively, and 1.357 for all older generations; the size at division for a particular cell was drawn from a narrow gaussian distribution near these mean values with coefficient of variation 0.15. The parameters of the simulations in units of average growth rate are: $k_r = 80$,

$\gamma_m = 4$, $k_g = 4$, $\gamma_g = 0$, $k_4 = 8$, $\gamma_4 = 0.5ga1^{-1}$, $K = 0.05$. The protocol of the simulations closely resembled the experimental procedure. We started from a small collection of cells (typically 10^3), with gaussian distributions of sizes and quantities of GAL4p, mRNA and GFP near the expected mean values. This small population was 'grown' until it reached 100,000 cells. The state of the population was recorded within a narrow time window right after that. We checked that this measurement was consistent with the time-ensemble average over a longer period of time, indicating that the evolution of the population reached the stationary state.

Received 3 June; accepted 29 September 2005.

Published online 21 December 2005.

1. Rao, C., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).
2. Reaney, D. Genetic noise in evolution? *Nature* **307**, 318–319 (2002).
3. Sternberg, P. W. & Felix, M. Evolution of cell lineage. *Curr. Opin. Genet. Dev.* **7**, 543–550 (1997).
4. Grossman, A. D. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu. Rev. Genet.* **29**, 477–508 (1995).
5. Kemkemer, R., Schrank, S., Vogel, W., Gruler, H. & Kaufmann, D. Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 *in vitro*. *Proc. Natl Acad. Sci. USA* **99**, 13783–13788 (2002).
6. Cook, D. L., Gerber, A. N. & Tapscott, S. J. Modeling stochastic gene expression: Implications for haploinsufficiency. *Proc. Natl Acad. Sci. USA* **95**, 15641 (1998).
7. Elowitz, M., Levine, A., Siggia, E. & Swain, P. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
8. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
9. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
10. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
11. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
12. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
13. Kepler, T. & Elston, T. Stochasticity in transcriptional regulation: origins, consequences and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
14. Hasty, J., Millen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
15. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
16. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100**, 7714–7719 (2003).
17. Kaern, M., Elston, T., Blake, W. & Collins, J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
18. Simpson, M. L., Cox, C. D. & Sayler, G. S. Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl Acad. Sci. USA* **100**, 4551–4556 (2003).
19. Hartwell, L. H. & Unger, M. Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *J. Cell Biol.* **75**, 422–435 (1977).
20. Woldringh, C. L., Huls, P. G. & Vischer, N. E. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/ α as determined by image cytometry. *J. Bacteriol.* **175**, 3174–3181 (1993).
21. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26**, 183–186 (2000).
22. Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**, 661–670 (2004).
23. Adalsteinsson, D., Millen, D. & Elston, T. Biochemical network stochastic simulator (bionets): Software for stochastic modeling of biochemical networks. *BMC Bioinformatics* **5**, 24 (2004).
24. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NSF Division of Molecular and Cellular Bioscience and the Alfred P. Sloan Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H. (hasty@ucsd.edu).

The primate amygdala represents the positive and negative value of visual stimuli during learning

Joseph J. Paton^{1*}, Marina A. Belova^{1*}, Sara E. Morrison¹ & C. Daniel Salzman^{1,2,3}

Visual stimuli can acquire positive or negative value through their association with rewards and punishments, a process called reinforcement learning. Although we now know a great deal about how the brain analyses visual information, we know little about how visual representations become linked with values. To study this process, we turned to the amygdala, a brain structure implicated in reinforcement learning^{1–5}. We recorded the activity of individual amygdala neurons in monkeys while abstract images acquired either positive or negative value through conditioning. After monkeys had learned the initial associations, we reversed image value assignments. We examined neural responses in relation to these reversals in order to estimate the relative contribution to neural activity of the sensory properties of images and their conditioned values. Here we show that changes in the values of images modulate neural activity, and that this modulation occurs rapidly enough to account for, and correlates with, monkeys' learning. Furthermore, distinct populations of neurons encode the positive and negative values of visual stimuli. Behavioural and physiological responses to visual stimuli may therefore be based in part on the plastic representation of value provided by the amygdala.

The complex anatomical connections of the amygdala, a collection of nuclei located deep in the medial temporal lobe, make it a prime candidate for providing a representation of the value of visual stimuli. The amygdala receives inputs from the visual system and from other sensory systems that represent reinforcing stimuli^{6–8}. In addition, the amygdala is likely to receive error signals that represent stimuli in relation to expectations and that may be essential in creating an updated representation of value. The source of error signals for aversive learning has not been identified; however, midbrain dopamine neurons might supply such error signals for appetitive learning⁹. These signals could influence amygdala neural responses either directly⁷ or indirectly through other brain structures such as the prefrontal cortex^{6,10}. The convergence of information from each of these input pathways, perhaps combined with processing that occurs through intrinsic connections within the amygdala, may form a representation of visual stimulus value.

Unlike the anatomy of the amygdala, the physiological properties of amygdala neurons—especially with respect to learning the value of visual stimuli—remain poorly understood. We therefore recorded the activity of single amygdala neurons while monkeys learned the positive or negative value of new, abstract images during a trace-conditioning procedure¹¹ (Fig. 1a). In this task, a brief time interval (the trace interval) separated the presentation of a conditioned stimulus (CS) from an unconditioned stimulus (US). In each experiment, we assigned every CS a value: positive, negative or non-reinforced. After the image viewing and trace intervals, we delivered a liquid reward (after positive images), an aversive air-puff

directed at the face (after negative images) or nothing (after non-reinforced images). The monkeys learned to associate each image with its value: during the trace interval they licked at a spout in anticipation of a liquid reward, or closed their eyes—a defensive behaviour—in anticipation of an air-puff. (We refer to the monkeys' eye closures as blinks, but the eyes typically stayed closed for longer than a blink, suggesting that this behaviour was purposeful.)

After the monkeys had learned the associations, without warning we reversed the values of the positive and negative images, and the monkeys learned the new values. Unlike a task in which subjects learn to avoid aversive USs, in our task, CSs predicted USs with 100% certainty. We tested every neuron during both appetitive and aversive conditioning for the same CS, allowing us to determine whether distinct amygdala neurons respond preferentially to positive or negative value for each CS. We report that amygdala neurons encode—in a flexible manner and in separate populations of neurons—the positive or negative value of conditioned images. Furthermore, these signals predict when monkeys learn, and develop rapidly enough to account for learning.

Anticipatory licking and blinking behaviour provided a measure of monkeys' learning about the positive and negative association of images. In the experiment shown in Fig. 2a, b, licking response rates were greater when an image was positive than when the same image was negative. Blinking responses showed the opposite trend. We used a change-point test to identify when the response rate changed significantly in relation to a value reversal ($P < 0.05$), represented graphically by a change in slope of the cumulative record of responses¹². For image 1, the licking rate decreased significantly at trial 52 and blinking rate increased at trial 48 (22 and 18 trials after reversal, respectively). Because the monkey still licked sporadically when image 2 was negative, the change-point algorithm identified trial 29 as when licking began to increase, slightly before the reversal. Blinking decreased significantly starting at trial 35.

The primary objectives of our study were (1) to determine whether positive and negative conditioned visual stimuli engage the same or different populations of amygdala neurons, (2) to compare quantitatively the time course of how neural activity and behaviour change upon image value changes, and (3) to characterize the relative contributions of conditioned value and image identity to neuronal responses. In each of two monkeys (monkey V and monkey P), we used magnetic resonance imaging (MRI) to localize our recording electrode to the amygdala (Fig. 1b–d). We obtained complete data sets during monkeys' performance of initial and reversal learning for 196 neurons. To determine the extent to which amygdala neural activity reflected the conditioned values of images, or the images' sensory properties, we examined neural activity in relation to image value reversals. For the amygdala neuron recorded during the behaviour depicted in Fig. 2a, b, neural activity during the trace

¹Center for Neurobiology and Behavior, ²W. M. Keck Center on Brain Plasticity and Cognition, and ³Department of Psychiatry, Columbia University, 1051 Riverside Drive, Unit 87, New York, New York 10032, USA.

*These authors contributed equally to this work.

interval was higher when images had a positive value than when the same images had a negative value (Fig. 2c–f), typical of a value-coding neuron. Other cells showed value-coding during the visual stimulus interval (see Supplementary Note 1 and Supplementary Fig. 1).

To characterize how quickly neural activity changed in relation to reversal of image value, we applied the change-point test¹² to trace interval neural responses (Fig. 2g, h). For image 1, neural activity began to decrease at trial 38 (8 trials after reversal), before either licking or blinking behaviour changed. For image 2, neural activity increased at trial 33, preceding the change in blinking behaviour. Overall, 100/196 (51%) of neurons changed activity in relation to value during the visual stimulus and/or trace intervals. For the cell shown in Fig. 2, a two-way analysis of variance (ANOVA) verified the influence of image value on neural responses (65% of variance accounted for by image value, $P < 0.05$), as well as a much smaller effect of image identity (8% of variance accounted for by image identity, $P < 0.05$). Across our population of neurons, image value often accounted for a statistically significant percentage of the variance (Fig. 3a, b; $P < 0.05$, two-way ANOVA). In addition, image identity had a significant effect in many cells, often overlapping with the representation of value at the single-cell level. The inset histograms show the difference in the strength of identity and value coding for all neurons with a significant effect of image identity, value or both factors. The two distributions are significantly different, with the trace interval showing significantly more value coding compared to image identity coding than the visual stimulus interval ($P < 10^{-5}$, Wilcoxon test; visual stimulus median difference -1.2 ; trace median difference 11.8). Thus, the representation of image identity weakens substantially after an image disappears, and neural activity then predominantly reflects the learned value of images.

To further characterize the representation of value in the amygdala,

we determined quantitatively whether neurons responded more strongly when an image had a positive or negative value. We used a receiver operating characteristic (ROC) analysis¹³ to estimate the extent to which activity before and after an identified change point differed (Fig. 3c). Each data point from this analysis reflects the change in activity for a single image in a single interval. By convention, we used the negative value responses as the reference distribution, so ROC values >0.5 indicated cells that responded more when an image was positive than when the same image was negative. All ROC values except two were significantly different from 0.5 ($P < 0.05$, permutation test). The bi-modally distributed data indicate that some amygdala neurons prefer negative conditioned stimuli, and others prefer positive conditioned stimuli (see Supplementary Note 2).

The activity of neurons encoding positive and negative values was not related to motor responses (licking and blinking; see Supplementary Figs 3, 4). Moreover, the development of value coding was not an artefact of using different modalities for our reinforcing stimuli (although the air-puff and liquid reward both have auditory and somatosensory components). Many neurons encoding positive image value responded to air-puffs, and many neurons whose activity reflected negative image value responded to rewards (Table 1 and Supplementary Fig. 5). Thus, neurons encoding value received information about both kinds of reinforcing stimuli. Finally, neurons representing positive and negative values did not demonstrate clear anatomical clustering (Fig. 1e, f).

We next characterized how neurons represent value across time during trials. An ROC analysis applied to consecutive overlapping time-windows of 100 ms (advanced in steps of 20 ms), revealed that different neurons encoded image value during different overlapping time segments (Fig. 3d). Some neurons represented value exclusively during the visual stimulus or trace intervals, but other cells sustained

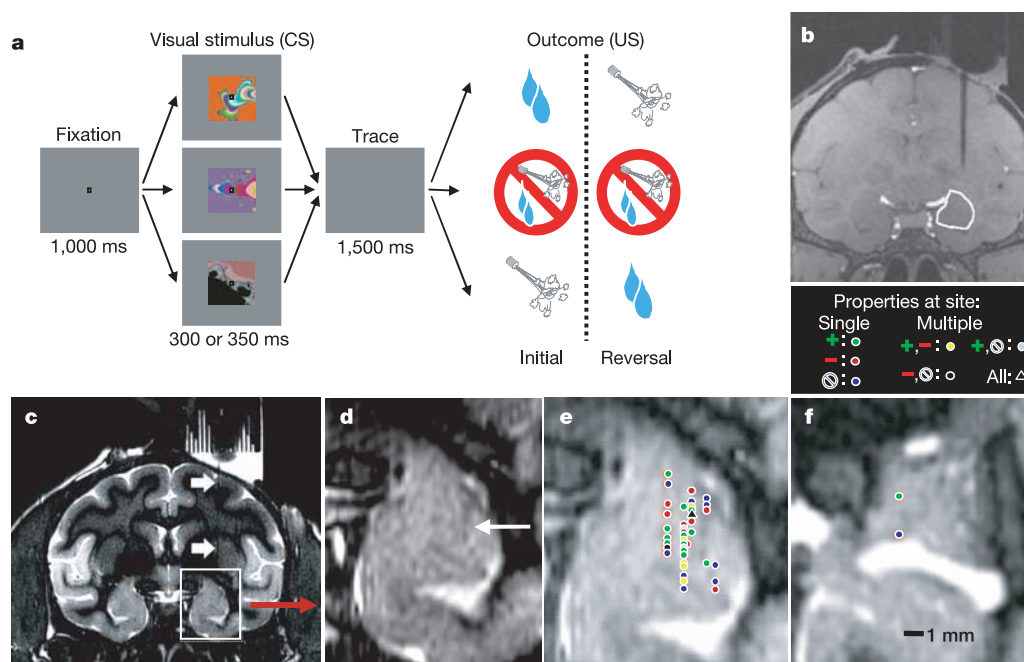


Figure 1 | Task and brain MRI. **a**, Sequence of events for the three trial types. Top and bottom squares, image values reverse. Middle square, image always non-reinforced. **b**, Coronal MRI acquired with a two-dimensional (2D) spoiled gradient recalled acquisition (SPGR) sequence in monkey V, showing the artefact from a tungsten microelectrode dorsal to the amygdala (outlined in white). **c–f**, Coronal MRI with 2D inversion recovery (IR) sequence (**c**, arrows point to the electrode). Magnified images show the probable border of the lateral nucleus (arrow in **d**) and recording site

locations (slice in **f** is immediately posterior to **e**). We collapsed recording sites spanning 2 mm in the anterior–posterior dimension onto each image slice, in many cases resulting in the superposition of multiple cells with different properties (see key above **f** for symbols denoting properties; ‘+’ denotes positive value-coding, ‘−’ denotes negative value-coding, and ‘no’ symbol indicates no value-coding). Recording sites from monkey P occurred in an overlapping region of the amygdala.

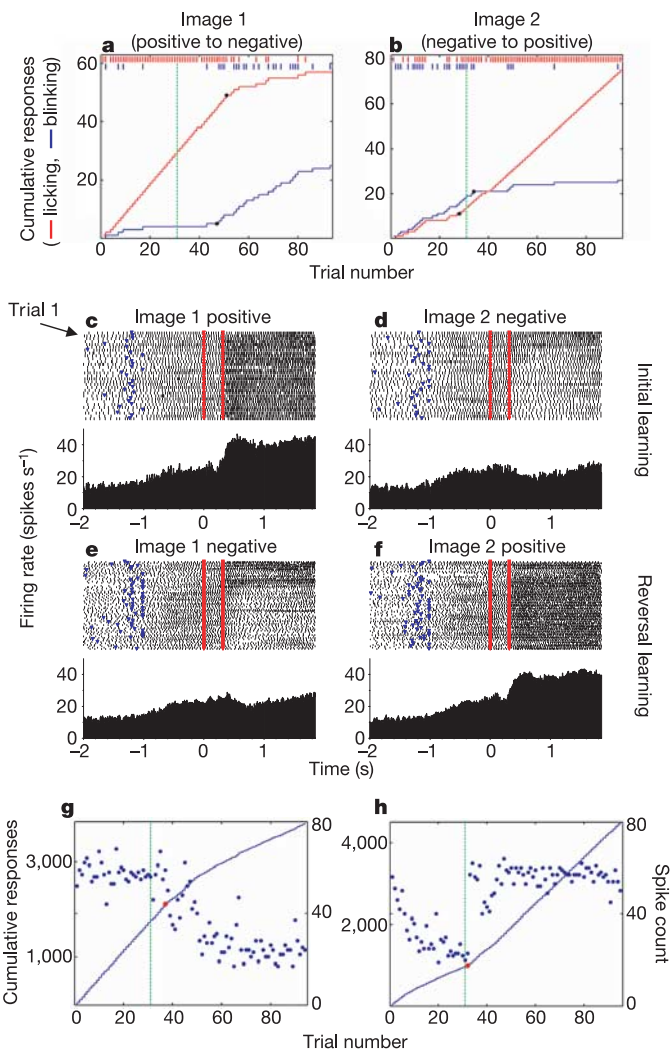


Figure 2 | Behaviour and neural activity from a single amygdala neuron during learning. **a, b,** Cumulative (curves) and trial-by-trial (tick marks) measures of licking (red) and blinking (blue), plotted as a function of trial number for images 1 and 2. Black dots indicate change points. Vertical green lines indicate value reversals. **c–f,** Rasters and peri-stimulus time histograms (PSTHs), truncated at US delivery, for the amygdala cell recorded during the same experiment. Each row of dots represents the timing of action potentials during one trial. PSTHs sum activity across trials and were smoothed by taking a 10-ms moving average of activity. Blue ticks indicate fixation point onset. Red ticks indicate visual stimulus onset/offset. **g, h,** Spike count and cumulative spike count during the trace interval, plotted as a function of trial number for images 1 and 2. Red dots indicate change points.

a representation across some or all of both intervals. Across the population of neurons, signals representing value developed rapidly after visual stimulus onset and were maintained through the trace interval (Fig. 3d; for positive-coding cells, the first bin significantly greater than 0.5 occurred 80–180 ms after visual stimulus onset,

$P < 0.05$ by t -test; for negative-coding cells, the first significant bin was 120–220 ms after visual stimulus onset). This temporally extended representation of value may be useful for the development of adaptive behaviour—such as licking and blinking—that anticipates reinforcement. Furthermore, the representation of value over time could have an important role in reinforcement learning: computations of error signals may depend on a comparison of neural signals that represent value as a function of time⁹.

If decisions to lick or blink in anticipation of a reward or punishment are based in part on the representation of value we describe in the amygdala, then, on average, neural activity should change at the same time as learning occurs. To compare behavioural learning of the image value reversals to changes in neural activity, we plotted the change-point for neural activity against the trial when behaviour changed (Fig. 4a, b; each data point compares a neural activity change point with either a blinking or licking change point; the distributions of licking and blinking change points were not significantly different from one another; $P > 0.25$, t -test). Learning was significantly correlated with changes in neural activity during both the visual stimulus and trace intervals (visual stimulus interval: monkey V, $P = 0.02$, $r^2 = 0.06$; monkey P, $P < 10^{-5}$, $r^2 = 0.44$; trace interval: monkey V, $P < 10^{-5}$, $r^2 = 0.40$; monkey P, $P < 10^{-5}$, $r^2 = 0.32$). The difference between neural and behavioural change points was near zero (Fig. 4 histograms; monkey V: visual stimulus mean difference 1.4 trials, trace mean difference -0.8 trials; monkey P: visual stimulus mean difference -0.5 trials, trace mean difference 0.3 trials). For all comparisons, the onsets of changes in behavioural and neural responses were statistically indistinguishable (paired t -test, $P > 0.05$). In addition, the distributions comparing change points during the visual stimulus and trace intervals were not significantly different ($P > 0.1$, t -test).

These results indicate that amygdala neurons can rapidly adjust their activity to reflect image value; these changes in activity both predict when monkeys learn and develop, on average, as fast as monkeys learn. The fact that some cells change their activity more rapidly than monkeys learn suggests that monkeys cannot selectively ‘listen’ to these neurons, or they would demonstrate faster learning. Learning may be better accounted for if monkeys must evaluate signals from the population of neurons, which includes neurons that change their activity more slowly.

The data presented so far compare the onset of changes in behavioural and neural responses. We next compared the time course of average neural and behavioural responses across the 100 neurons encoding image value (as identified by the change-point test). We normalized and then averaged neural activity and behaviour from the 20 trials before and after the value reversal of each image. To describe the behavioural and neural changes around reversals, we fitted the behavioural and neural data with sigmoid functions (Fig. 4c, see Supplementary Fig. 6). The 95% prediction intervals for these functions overlapped, indicating that the trajectories of changes in behavioural responses and neural activity were similar under this model. In contrast, the same analysis on the non-reinforced images for the same cells showed no change in activity or behaviour in relation to the reversal of the positive and negative images (Fig. 4d). These data demonstrate that, on average, neurons begin to change their activity within very few trials of a change in image value, and

Table 1 | Responses to reinforcing stimuli in amygdala neurons that code value

	Excitatory response to air-puff	Inhibitory response to air-puff	Excitatory response to reward	Inhibitory response to reward	No response to reward or air-puff
Cells coding positive CS value (n = 61)	34	12	23	6	9
Cells coding negative CS value (n = 39)	18	8	11	7	7

Excitatory and inhibitory responses to air-puffs and rewards for all cells encoding the positive or negative value of images. The number of cells adds up to more than the number of value-coding neurons because many cells were modulated by both air-puff and reward.

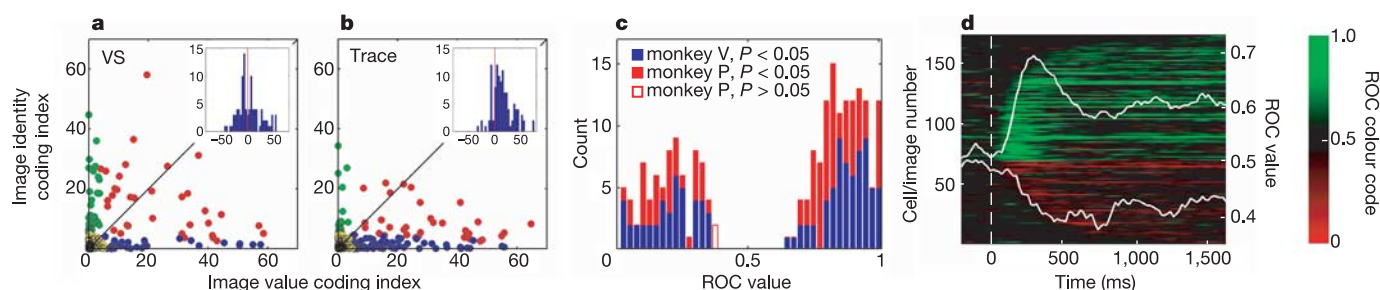


Figure 3 | Amygdala neurons encode the positive and negative value of visual stimuli. **a, b,** Image identity coding index (II_CI) plotted against image value coding index (IV_CI) for the visual stimulus (VS) and trace intervals. Green, blue and red dots, $P < 0.05$ for II_CI, IV_CI or both indices, respectively (two-way ANOVA). Yellow dots, not significant. Inset histograms show II_CI subtracted from IV_CI for non-yellow data points. See Supplementary Fig. 9. **c,** Encoding of positive and negative value shown by ROC analysis. **d,** Development of value signals as a function of time. Each

row in the colour map represents value coding for a neuron during presentation of a single image, with positive and negative cell rows sorted in opposite order according to the first post-visual-stimulus data point significantly different from 0.5 ($P < 0.05$, permutation test). White curves show mean ROC values across the positive- and negative-coding populations. Time 0 is the start of the bin spanning from 0–99 ms after visual stimulus onset.

that—across the population of neurons—the rate of changes in activity closely tracks behavioural learning.

The rapid appearance of a representation of visual stimulus value in the amygdala raises the possibility that this representation is formed locally. If a representation of value is first formed elsewhere in the brain, it would have to provide signals that represent value at earlier time points within trials, and activity would have to change even more rapidly across trials compared to what we observe in the amygdala. Moreover, the correlation between changes in amygdala activity and behaviour strongly suggests that the amygdala representation is used by the monkey, regardless of its source. Nonetheless, other brain structures such as the orbitofrontal cortex probably have critical roles in learning on our task. The amygdala seems to update its representation as quickly as the orbitofrontal cortex changes its

activity during reversal learning (using a different task, in which monkeys learn to avoid aversive stimuli, and different analytic methods)¹⁴. Other studies that have not used aversive stimuli also suggest that the orbitofrontal cortex provides a representation of reward value^{15–17}. Understanding the complex interactions between the amygdala and orbitofrontal cortex (and other structures) will require substantial investigative effort in order to tease apart the distinct contributions of each brain area to reinforcement learning. In rodents, some evidence suggests that the physiological properties of amygdala and orbitofrontal neurons are interdependent^{18,19}.

Human and non-human primate behaviour is strongly influenced by visual processing and, therefore, the assignment of value to visual stimuli is a fundamental process that underlies many of our actions. In lower mammals, previous studies have investigated amygdala neurophysiology in relation to learning; however, these studies have used auditory or olfactory conditioned stimuli, and tasks that either do not include appetitive conditioning or that involve learning to avoid an aversive unconditioned stimulus^{1,20}. The primate amygdala has extensive interconnections with the visual cortex²¹, and previous lesion studies in monkeys and neuroimaging studies in humans have suggested that the amygdala is involved in updating a representation of the value of visual stimuli^{3,22,23} (see Supplementary Note 3). Previous neurophysiological recordings in monkey amygdala have used different methodological procedures than those here, and provide conflicting data regarding whether the amygdala contains an updated representation of visual stimulus value^{24–26}.

Our data establish that distinct amygdala neurons encode the positive and negative learned value of images. The representation of value is often linked to image selectivity initially, but this representation becomes transformed over time so that value, and not image selectivity, is predominantly encoded after visual stimuli disappear. Signals that reflect value develop rapidly, and evolve as quickly, on average, as monkeys learn. Furthermore, the timing of changes in amygdala activity predicts when monkeys learn. Thus, in principle, decisions to either lick or blink during performance of our task could be based on the plastic representation of visual stimulus value provided by the amygdala.

Reinforcement learning, a process by which sensory stimuli become associated with positive or negative values, is intimately related to physiological and behavioural responses thought to represent emotions, because such responses frequently occur as a reaction to sensory stimuli endowed with value. Amygdala neural signals that reflect CS value may influence a variety of interconnected brain structures—including the ventral striatum, prefrontal cortex, sensory cortices, the hippocampus and rhinal cortices, and subcortical brain areas—involved in aspects of emotional responses and reinforcement learning. Different amygdala nuclei have complex

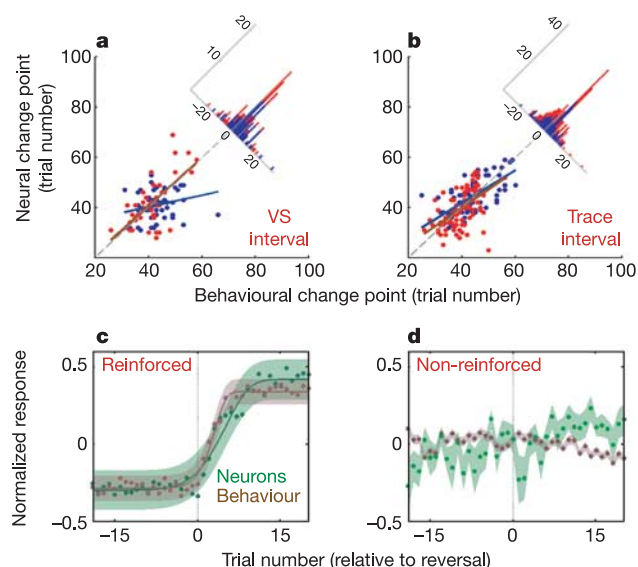


Figure 4 | The relationship between changes in neural activity and behavioural responses. **a, b,** Onset of changes in neural activity plotted against the onset of changes in licking or blinking for visual stimulus (VS) and trace interval activity. Histograms show the difference between neural and behavioural change points. Blue, data and regression lines for monkey V; red, data and regression lines for monkey P. **c,** Average normalized neural activity and behavioural responses plotted as a function of trial number relative to the reversal in image value. Shaded regions indicate 95% prediction intervals for best-fit Weibull functions. **d,** Similar representation as in **c**, applied to the non-reinforced images. Shaded regions show s.e.m. of data points.

patterns of connectivity with each of these structures^{6,7,10,21,27}, suggesting that neural signals in the amygdala may be used for very different purposes depending upon their anatomical destination. For example, signals that reflect CS value may influence attention²⁸ or computations of error signals⁹ through projections from the amygdala central nucleus to the ventral tegmental area and substantia nigra²⁹. In addition, value signals could influence behavioural and physiological responses to visual stimuli through other subcortical projections from the central nucleus³⁰, and through projections from the basal and accessory basal nuclei to the prefrontal cortex and ventral striatum^{6,7,10}. The amygdala neural signals we describe—which reflect the learned value of visual stimuli—are therefore only one component of the intricate neural circuitry that mediates reinforcement learning and our emotional lives. To understand how this circuitry operates, future studies must identify how dynamic interactions between the amygdala and interconnected brain structures function to support a wide range of emotional and cognitive behaviours.

METHODS

Behavioural task. In every experiment, monkeys learned the positive, negative or non-reinforced value of three new images through a trace-conditioning procedure. In individual trials, monkeys centred their gaze at a fixation point (FP) for 1 s and viewed an image for 300 or 350 ms, determined empirically for each monkey to minimize broken fixation behaviour (see Supplementary Fig. 7). After a 1,500-ms trace interval during which monkeys were free to move their eyes, we delivered a US (for positive images, a 0.2–0.9 ml liquid reward; for negative images, a 50–100 ms, 40–60 psi aversive air-puff directed at the face; for non-reinforced images, nothing). We reversed image value assignments without warning at least 8 trials after monkeys learned the initial contingencies, as assessed by viewing licking and blinking data during the experiment. All trials were pseudorandomly interleaved. Images were fractal patterns constructed with a custom-written Matlab (Mathworks) program. Images had no inherent value and were easily distinguishable from each other and from recently shown images on previous days.

Data collection. We recorded neural activity from the right amygdala of two rhesus monkeys (*Macaca mulatta*) weighing 4–10 kg. All animal procedures conformed to NIH guidelines and were approved by the Institutional Animal Care and Use Committees at New York State Psychiatric Institute and Columbia University. We positioned recording chambers based on anatomical information acquired using MRI. Conventional metal microelectrodes (1–4) (FHC Instruments) were individually controlled and advanced into the brain through dura-puncturing guide tubes using either a motor-controlled hydraulic microdrive (Narishige) or a motorized multi-electrode drive (NAN). Signal amplification, filtering, digitizing of spike waveforms, and spike-sorting using a principal component analysis platform (online, with offline verification) were accomplished using the Plexon system. While we searched for neurons, monkeys performed either no task or a fixation task without images present. We studied all well-isolated neurons (196 neurons in total: 92 from monkey V, 104 from monkey P). We reconstructed recording sites using MRI combined with laboratory notes. Neurons were located in the middle 50% of the anterior–posterior extent of the amygdala, in the lateral 75% of the medial–lateral extent, and in the dorsal 75% of the dorsal–ventral extent. All neurons were probably located in either the basal, accessory basal, lateral or central nucleus, or the intercalated masses.

Licking was measured by the interruption of an infrared beam by the tongue extending to the reward tube, and blinking by the loss in signal from an infrared eye tracker (ASL), which corresponded to blinks as verified by a video camera feed. Licks or blinks occurring during the 500 ms preceding US delivery were scored as responses.

Data analysis. Based on an analysis of visual response latency (see Supplementary Fig. 8), we divided the trial into visual stimulus (90 ms after image appearance until 90 ms after image disappearance) and trace (90 ms after image disappearance until US delivery) intervals. Our data analysis had four principal goals. (1) To find the trial closest to image value reversals at which neural and behavioural (licking, blinking) responses began to change, we used a change-point test¹² ($P < 0.05$, correcting for multiple comparisons). (2) To characterize the relative contribution of image identity (that is, the sensory properties of images) and value on neural activity, we performed a two-way ANOVA. We defined three indices, corresponding to the percentage of variance accounted for by image identity, image value and their interaction (image identity, image value, and image value–identity interaction coding indices). (3)

To determine whether amygdala neurons responded more when an image had a positive or negative value, we compared spike counts from the 20 trials on either side of an identified change point using a receiver operating characteristic (ROC) analysis¹³ on neurons categorized as value-coding (209 intervals from 100 cells coding image value during the visual stimulus, trace or both intervals; 72/100 cells changed activity in opposite directions for both images upon reversal; 28/100 neurons changed responses for one image, but two-way ANOVA showed significant effects of image value). We also characterized the development within trials of signals representing value, computing an ROC value in 100-ms windows moved in 20-ms steps across the trial (performed on 172 image/cell combinations from the 100 value-coding cells). (4) To compare the trajectories of changes in neural and behavioural responses, we normalized and averaged neural activity and behaviour from each of the 20 trials before and after image value reversal across the 100 value-coding neurons. We fitted the behavioural and neural data with a Weibull function:

$$f(x) = u + (1 - u) \exp(-x/\alpha)^\beta \quad (1)$$

which modelled normalized response as a function of trial number. We performed a similar normalization and averaging procedure for non-reinforced trials, analysing the 20 trials before and after the trial that the other images reversed. Details of all methods are described further in Supplementary Methods.

Received 20 September; accepted 25 November 2005.

- Maren, S. & Quirk, G. J. Neuronal signaling of fear memory. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Baxter, M. & Murray, E. A. The amygdala and reward. *Nature Rev. Neurosci.* **3**, 563–573 (2002).
- Everitt, B. J., Cardinal, R. N., Parkinson, J. A. & Robbins, T. W. Appetitive behaviour: impact of amygdala-dependent mechanisms of emotional learning. *Ann. NY Acad. Sci.* **985**, 233–250 (2003).
- Holland, P. C. & Gallagher, M. Amygdala circuitry in attentional and representational processes. *Trends Cogn. Sci.* **3**, 65–73 (1999).
- Stefanacci, L. & Amaral, D. G. Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. *J. Comp. Neurol.* **451**, 301–323 (2002).
- Amaral, D., Price, J., Pitkanen, A. & Carmichael, S. in *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction* (ed. Aggleton, J.) 1–66 (Wiley-Liss, New York, 1992).
- McDonald, A. J. Cortical pathways to the mammalian amygdala. *Prog. Neurobiol.* **55**, 257–332 (1998).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Ghashghaie, H. T. & Barbas, H. Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* **115**, 1261–1279 (2002).
- Pavlov, I. P. *Conditioned Reflexes* (Oxford Univ. Press, London, 1927).
- Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proc. Natl Acad. Sci. USA* **101**, 13124–13131 (2004).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (John Wiley and Sons, New York, 1966).
- Rolls, E. T., Critchley, H. D., Mason, R. & Wakeman, E. A. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J. Neurophysiol.* **75**, 1970–1981 (1996).
- Wallis, J. D. & Miller, E. K. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* **18**, 2069–2081 (2003).
- Roesch, M. R. & Olson, C. R. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304**, 307–310 (2004).
- Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
- Saddoris, M. P., Gallagher, M. & Schoenbaum, G. Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. *Neuron* **46**, 321–331 (2005).
- Schoenbaum, G., Setlow, B., Saddoris, M. P. & Gallagher, M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39**, 855–867 (2003).
- Schoenbaum, G., Chiba, A. & Gallagher, M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neurosci.* **1**, 155–159 (1998).
- Freese, J. L. & Amaral, D. G. The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *J. Comp. Neurol.* **486**, 295–317 (2005).
- Gottfried, J., O'Doherty, J. & Dolan, R. J. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* **301**, 1104–1107 (2003).
- LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E. & Phelps, E. A. Human

- amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* **20**, 937–945 (1998).
24. Sanghera, M. K., Rolls, E. T. & Roper-Hall, A. Visual responses of neurons in the dorsolateral amygdala of the alert monkey. *Exp. Neurol.* **63**, 610–626 (1979).
 25. Rolls, E. in *The Amygdala: A Functional Analysis* (ed. Aggleton, J.) 447–478 (Oxford Univ. Press, New York, 2000).
 26. Nishijo, H., Ono, T. & Nishino, H. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J. Neurosci.* **8**, 3570–3583 (1988).
 27. Stefanacci, L., Suzuki, W. A. & Amaral, D. G. Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. *J. Comp. Neurol.* **375**, 552–582 (1996).
 28. Lee, H. J. *et al.* Role of amygdalo-nigral circuitry in conditioning of a visual stimulus paired with food. *J. Neurosci.* **25**, 3881–3888 (2005).
 29. Fudge, J. L. & Haber, S. N. The central nucleus of the amygdala projection to dopamine subpopulations in primates. *Neuroscience* **97**, 479–494 (2000).
 30. Davis, M. in *The Amygdala: A Functional Analysis* (ed. Aggleton, J.) 213–287 (Oxford Univ. Press, New York, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. R. Gallistel for discussions and for assistance with the change-point test; S. Dashnaw and J. Hirsch for MRI support; C. A. Mason, E. R. Kandel, M. N. Shadlen and members of the Mahoney Center at Columbia for comments on the manuscript; and M. E. Goldberg, J. E. LeDoux and W. T. Newsome for mentoring during a career development award to C.D.S. This work was supported by the Keck foundation, grants from the NIMH, and the Klingenstein, Sloan and NARSAD foundations, and by a Charles E. Culpeper Scholarship award from Goldman Philanthropic Partnerships to C.D.S. J.J.P. received support from NICHD and NEI institutional training grants. S.E.M. received support from an NSF graduate research fellowship.

Author Contributions J.J.P. and M.A.B. performed all experiments and conducted data analyses. S.E.M. performed some of the data analyses and contributed to many discussions. Experiments were designed and implemented in the laboratory of C.D.S.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D.S. (cds2005@columbia.edu).

The Polycomb group protein EZH2 directly controls DNA methylation

Emmanuelle Viré¹, Carmen Brenner¹, Rachel Deplus¹, Loïc Blanchon¹, Mario Fraga², Céline Didelot¹, Lluís Morey³, Aleyde Van Eynde⁴, David Bernard¹, Jean-Marie Vanderwinden⁵, Mathieu Bollen⁴, Manel Esteller², Luciano Di Croce³, Yvan de Launoit^{1,6} & François Fuks¹

The establishment and maintenance of epigenetic gene silencing is fundamental to cell determination and function¹. The essential epigenetic systems involved in heritable repression of gene activity are the Polycomb group (PcG) proteins^{2,3} and the DNA methylation^{4,5} systems. Here we show that the corresponding silencing pathways are mechanistically linked. We find that the PcG protein EZH2 (Enhancer of Zeste homolog 2) interacts—within the context of the Polycomb repressive complexes 2 and 3 (PRC2/3)—with DNA methyltransferases (DNMTs) and associates with DNMT activity *in vivo*. Chromatin immunoprecipitations indicate that binding of DNMTs to several EZH2-repressed genes depends on the presence of EZH2. Furthermore, we show by bisulphite genomic sequencing that EZH2 is required for DNA methylation of EZH2-target promoters. Our results suggest that EZH2 serves as a recruitment platform for DNA methyltransferases, thus highlighting a previously unrecognized direct connection between two key epigenetic repression systems.

The PcG protein EZH2 is a histone methyltransferase associated with transcriptional repression. EZH2 catalyses the addition of methyl groups, at least partly, to histone H3 at lysine 27 (H3K27)^{6–9}. Methylated K27 serves as an anchorage point for the recruitment of additional PcG proteins^{6–8}, the binding of which contributes to formation of a repressive chromatin state. Besides EZH2, DNMTs are also key epigenetic regulators involved in transcriptional repression^{10–12}. In the work presented here, we examined whether EZH2 and DNMTs might function through a common mechanistic pathway to silence gene expression. As shown in Fig. 1a, DNA methyltransferase activity was purified from HeLa nuclear extracts in ‘pull-down’ experiments using EZH2 fused to glutathione S-transferase (GST) (Fig. 1a, lane 3). The association of EZH2 with DNMT enzymatic activity was specific, since only background activity was measured in control experiments performed with GST alone, or with another GST-fused protein (PLZF, for promyelocytic leukaemia zinc finger) (Fig. 1a, lanes 1 and 2, respectively). GST–SUV39H1 (a H3K9 methyltransferase known to interact with DNMT¹³) was used as a positive control for the assay (Fig. 1a, lane 4). Furthermore, an antibody raised against EZH2 specifically precipitated DNA methyltransferase activity from nuclear extracts of untransfected HeLa cells (Fig. 1b, lane 3).

One or more among several DNA methyltransferases might be responsible for the EZH2-bound DNMT activity. To address this issue, we performed GST pull-downs with GST–EZH2, incubated with *in vitro* translated (IVT) ³⁵S-labelled DNMT1, DNMT3A or DNMT3B. As shown in Fig. 2a, all three DNMTs were able to bind to full-length GST–EZH2 (Fig. 2a, lane 3), in contrast to GST alone

(Fig. 2a, lane 2). GST–SUV39H1 was again used as a positive control (Fig. 2a, lane 7). Using various GST-fused EZH2 fragments, we found that EZH2 associates with DNMTs primarily via its amino-terminal portion (Fig. 2a, lanes 4–6). In the reciprocal experiment, IVT EZH2 was found to interact with both the amino-terminal and carboxy-terminal parts of DNMT1, and with the conserved PHD (plant homeodomain)-like motif in DNMT3A and DNMT3B (Supplementary Fig. S1). To demonstrate an *in vivo* interaction between EZH2 and DNMTs, we performed co-immunoprecipitations from extracts of untransfected cells. Endogenous EZH2 co-immunoprecipitated with endogenous DNMT1, DNMT3A and DNMT3B (Fig. 2b). The interaction of EZH2 with DNMTs was further confirmed by reverse endogenous co-immunoprecipitation of DNMTs with EZH2 (Fig. 2c, lane 2). EZH2 is known to associate with other PcG proteins, EED and SUZ12, within the context of the PRC2/3 complexes¹⁴. In GST pull-down experiments, we found that DNMTs can interact with EED and that endogenous EED and SUZ12 both associate with DNMT enzymatic activity (Supplementary Fig. S2). EED also co-immunoprecipitated with DNMTs from extracts of untransfected cells (Fig. 2c, lane 3). Taken together, these results indicate that native EZH2 interacts with DNMTs in the context of the PRC2/3 complexes and associates with DNA methyltransferase activity *in vivo*.

Next, we probed the functional relationship between EZH2 and DNA methyltransferases. As these proteins act as transcriptional repressors^{11,12,15}, we investigated whether they can silence a common target gene. Recent work has identified several EZH2-target genes, including the *MYT1* gene¹⁶. We first examined whether *MYT1* could be reactivated by addition of the DNA methylation inhibitor 5-aza-deoxycytidine (5-aza-dC). Measuring levels of messenger RNA in U2OS cells by reverse transcription followed by polymerase chain reaction analysis (RT–PCR), we observed elevated transcript levels of *MYT1* after cell treatment with 5-aza-dC (Fig. 3a). In contrast, 5-aza-dC treatment did not affect the expression of an unrelated house-keeping gene (*actin*) (Fig. 3a). Thus, the EZH2-regulated gene *MYT1* is specifically reactivated by 5-aza-dC.

We next investigated whether DNA methyltransferases might contribute with EZH2 to *MYT1* silencing. For this, we performed RNA interference (RNAi) experiments in U2OS cells to knock down EZH2, DNMT1, DNMT3A or DNMT3B (Supplementary Fig. S3). As illustrated in Fig. 3b (lane 2), *MYT1* mRNA levels were markedly elevated in cells with reduced EZH2 levels (RNAi EZH2), in keeping with previous observations¹⁶. Similarly, knockdown of any one of the three DNMTs resulted in higher *MYT1* expression (Fig. 3b, lanes 3–5). These data indicate that silencing of *MYT1* gene expression requires both EZH2 and DNA methyltransferases.

¹Free University of Brussels, Faculty of Medicine, Laboratory of Molecular Virology, 808 route de Lennik, 1070 Brussels, Belgium. ²CNIO, Cancer Epigenetics Group, C/- Melchor Fernández Almagro 3, 28029 Madrid, Spain. ³ICREA and CRG, Passeig Marítim 37–49, E-08003 Barcelona, Spain. ⁴Division of Biochemistry, Faculteit Geneeskunde, Katholieke Universiteit Leuven, Herestraat 49, 3000 Leuven, Belgium. ⁵Free University of Brussels, Faculty of Medicine, Laboratory of Neurophysiology, 808 route de Lennik, 1070 Brussels, Belgium. ⁶UMR 8117, CNRS, Institut Pasteur de Lille, Université de Lille 1, Institut de Biologie de Lille, 1 rue Calmette, 59021 Lille, Cedex, France.

The results described above suggest that both EZH2 and DNMTs might bind to the *MYT1* promoter. To test this possibility, we performed chromatin immunoprecipitation (ChIP) experiments. As shown in Fig. 4a (left panel, lane 1, 'RNAi control'), we observed binding of EZH2 to the *MYT1* promoter. DNMT1, DNMT3A and DNMT3B were also found to bind to this promoter (Fig. 4a, left panel, lane 1, 'RNAi control'). Thus, both EZH2 and DNMTs associate with the *MYT1* promoter.

Next, we tested whether EZH2 is required for DNMT binding to the *MYT1* promoter using the RNAi EZH2 cells. In control experiments, EZH2 depletion resulted, as expected, in reduced EZH2

binding, concomitant with reduced H3K27 trimethylation (H3K27me3)(Fig. 4a, left panel, compare lane 1 with lane 2). More importantly, the decrease in EZH2 and H3K27me3 occupancy at the promoter coincided with a significant reduction of DNMT1, DNMT3A and DNMT3B binding to *MYT1* (Fig. 4a, left panel, compare lane 1 with lane 2). It is noteworthy that RNA polymerase II (RNA pol II) was detected at the *MYT1* promoter after EZH2 knockdown (Fig. 4a, left panel, lane 2), but not when EZH2 and DNMTs were bound to the promoter (Fig. 4a, left panel, lane 1). This is consistent with the reactivation of *MYT1* gene expression, highlighted in Fig. 3. These findings can be extended to additional known EZH2-target promoters—*WNT1*, *KCNA1* and *CNR1* (Fig. 4a, right panel; ref 16): in keeping with our observations with *MYT1*, we found that binding of DNMTs to these promoters is also

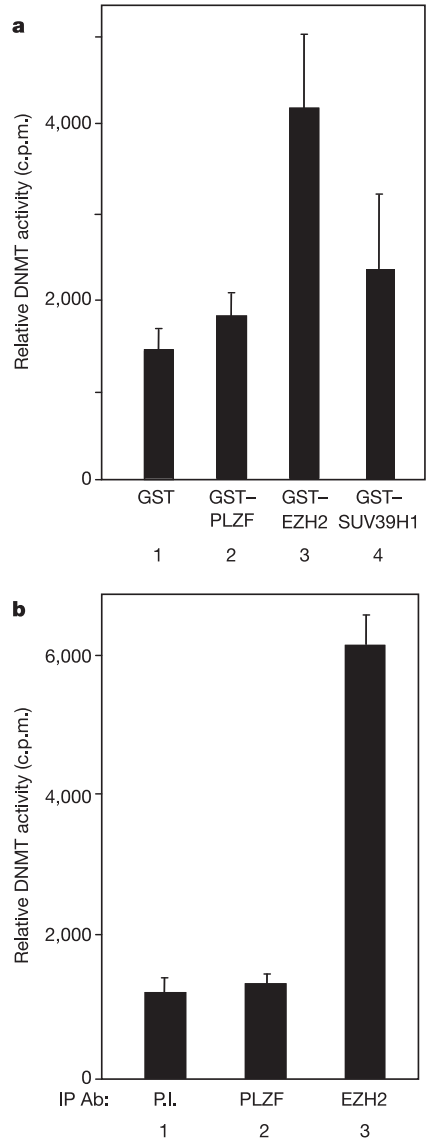


Figure 1 | The EZH2 Polycomb group protein associates with DNA methyltransferase activity. **a**, A GST-fused EZH2 protein was used to purify DNA methyltransferase activity from nuclear extracts. GST-SUV39H1 and GST-PLZF were also used as a positive¹³ and a negative²⁷ control, respectively. After incubation, the beads were washed and assayed for DNA methyltransferase activity read as c.p.m. of S-adenosyl-L[methyl-³H]methionine incorporated into an oligonucleotide substrate¹¹. **b**, Endogenous EZH2 associates with DNA methyltransferase activity. HeLa nuclear extracts were immunoprecipitated (IP) with either pre-immune serum (P.I.), a control antibody (ab) against PLZF²⁷, or an antibody against EZH2, and tested for associated DNA methyltransferase activity. The results are averages of at least three independent experiments with error bars showing standard deviations (s.d.).

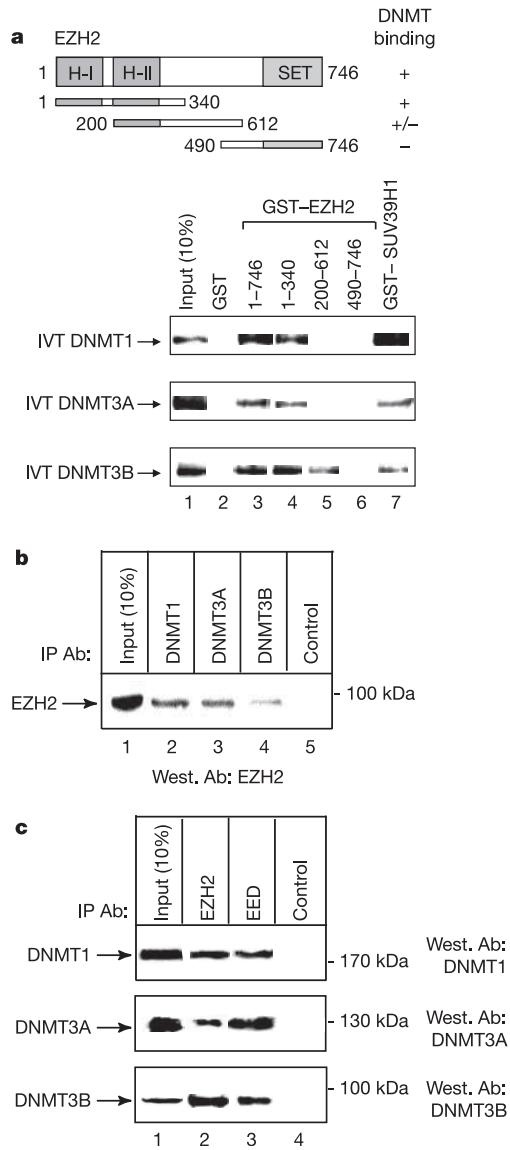


Figure 2 | EZH2 interacts with DNA methyltransferases in vitro and in vivo. **a**, *In vitro* translated (IVT) full-length DNMT1, DNMT3A or DNMT3B was incubated with the indicated GST fusions of EZH2. GST-SUV39H1 was used as a positive control for association with IVT DNMT proteins. **b**, EZH2 co-immunoprecipitates with DNA methyltransferases from untransfected cells. HeLa nuclear extracts were immunoprecipitated with the indicated antibodies. The presence of EZH2 in immunoprecipitates was visualized by western (west.) blot analysis using an antibody against EZH2. **c**, Endogenous co-immunoprecipitations of DNMTs with EZH2 and EED.

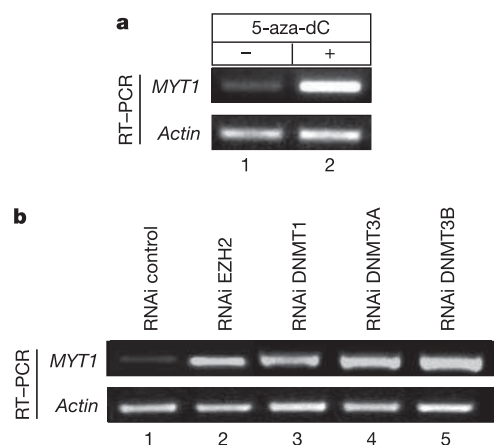


Figure 3 | Both EZH2 and DNA methyltransferases are required to repress the endogenous *MYT1* gene. **a**, The DNA methylation inhibitor 5-aza-dC reactivates *MYT1* expression. U2OS cells were either treated (lane 2) or not treated (lane 1) with 5-aza-dC (1 μ M). RT-PCR was performed to detect the presence of *MYT1* and *actin* transcripts. **b**, EZH2 or DNA methyltransferase depletion resulted in de-repression of endogenous *MYT1* expression. RT-PCR analysis of *MYT1* expression was performed on control cells and on the indicated single-gene knockdown cells.

compromised when association of EZH2 with the promoters is reduced (Fig. 4a, right panel). Together, these data indicate that the presence of EZH2 is necessary for binding of DNA methyltransferases to EZH2-target promoters. In contrast to these observations, DNA methyltransferases do not seem to be required for EZH2 or H3K27me3 binding to *MYT1* (Supplementary Fig. S4).

Having shown that EZH2 interacts physically with the DNA methyltransferases and facilitates their binding to EZH2-target promoters, we next examined whether EZH2 might influence their DNA methylation status. For this, we carried out bisulphite sequencing on *MYT1* and *WNT1* in EZH2-overexpressed cells or in EZH2-knockdown cells. Overexpression of wild-type EZH2 was

found to increase CpG methylation of the promoters significantly, whereas overexpression of an EZH2 mutant, lacking the conserved SET domain (for Suppressor of variegation 3–9, Enhancer of Zeste, Trithorax), failed to induce DNA methylation (Supplementary Fig. S5). As shown in Fig. 4b, we found that a number of CpGs located within *MYT1* or *WNT1* were significantly less methylated in EZH2-depleted cells than in RNAi control cells (CpGs situated outside the regions presented showed no significant difference in DNA methylation status; data not shown). Taken together, these results indicate that EZH2 is needed for CpG methylation of EZH2-target promoters.

The results presented here reveal that the Polycomb group protein EZH2 controls CpG methylation through direct physical contact with DNA methyltransferases (Fig. 4c). It is well established that in various organisms H3K9 methylation can be a prerequisite to DNA methylation^{17–19}, which in mammals involves the SUV39H enzymes and also probably the HP1 (heterochromatin protein 1) adaptor^{13,19}. It should be noted, however, that there are situations in which H3K9 and H3K27 methylation silence gene expression without the involvement of DNA methylation^{20,21}. Nevertheless, our finding that EZH2-mediated H3K27 methylation can dictate CpG methylation adds complexity and versatility to the emerging picture in which methylation of histones and DNA can work hand-in-hand as part of an epigenetic program integrating gene-silencing networks within the cell. Our finding that the Polycomb group protein EZH2 associates directly with DNA methyltransferases may be relevant to many different processes. It may relate to the *de novo* methylation activity of DNMT3 proteins involved in regulating gene expression^{10,12}. In addition, it may involve the maintenance functions which all three DNMTs possess^{22,23} and which may be necessary for the stable inheritance of silenced epigenetic states. More precisely, our discovery suggests a possible mechanism for epigenetic memory, in which an ‘epigenetic memory module’ consisting of both EZH2 and DNMTs might allow coordinated and heritable transmission of silenced epigenetic states through DNA replication. The close connection identified here between EZH2 and DNA methyltransferases might also be relevant to X-chromosome inactivation in female mammals, where both EZH2 and DNMTs are known to contribute

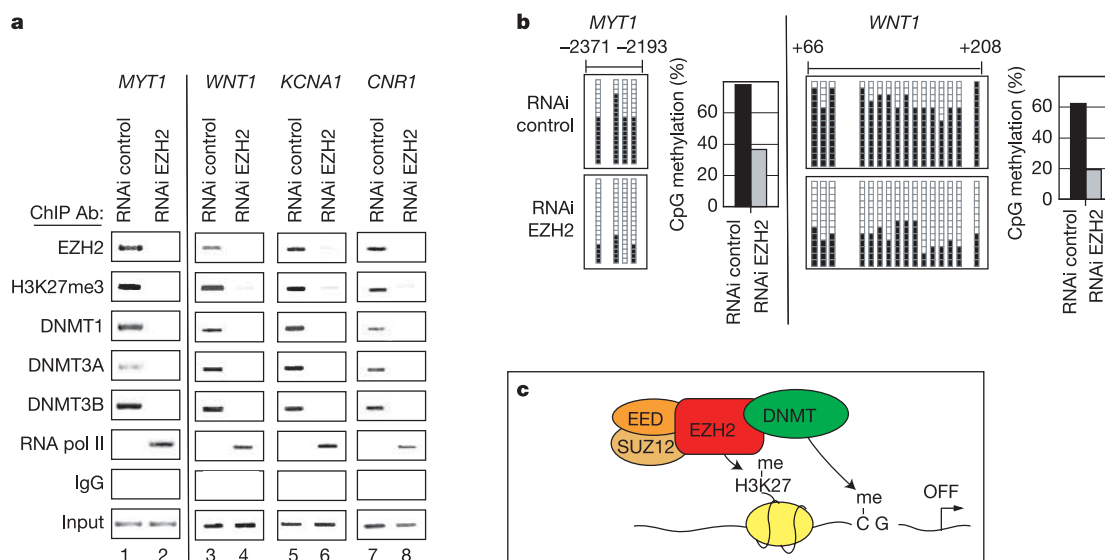


Figure 4 | EZH2 is required for binding of DNA methyltransferases and facilitates CpG methylation of EZH2-target promoters. **a**, ChIPs show that association of DNA methyltransferases with *MYT1* (left panels), as well as with other known EZH2-target genes (right panels; *WNT1*, *KCNA1* and *CNR1*; ref. 16), depends on EZH2. Cross-linked chromatin from mock-knockdown U2OS cells (RNAi control, lane 1) or EZH2-knockdown cells (RNAi EZH2, lane 2) was immunoprecipitated with the indicated

antibodies. **b**, Bisulphite sequencing shows that depletion of EZH2 (RNAi EZH2) leads to decreased methylation at a number of CpG sites within the *MYT1* (left panel) or the *WNT1* (right panel) promoter. Each row represents a single sequence analysed, and each square is a single CpG site. White and black squares represent unmethylated and methylated CpGs, respectively. **c**, EZH2 controls CpG methylation, in the context of the PRC2/3 complexes, through direct physical contact with DNA methyltransferases.

to maintaining the inactive state^{24–26}. Taken together, our results provide the first glimpse of a mechanistic link between two essential epigenetic systems: the Polycomb group proteins and the DNA methylation systems.

METHODS

GST fusion, pull-down assays and DNA methyltransferase assays. GST pull-downs and DNA methyltransferase assays were carried out as previously described^{11,27}. Plasmid constructs and antibodies used are detailed in the Supplementary Methods.

Immunoprecipitations and western blot analysis. Standard procedures were used for endogenous immunoprecipitations and western blotting¹³. See the Supplementary Methods for antibodies used.

Cell cultures and RNA interference. U2OS cells and the packaging cell line 293 GP were maintained in Dulbecco's modified Eagle's medium (DMEM, Gibco-BRL) supplemented with 8% fetal bovine serum and grown at 37 °C under 5% CO₂. See the Supplementary Methods for RNA interference.

RNA purification and RT–PCR analysis. RT–PCR was performed as described previously²⁷. Primer sequences are available on request.

Chromatin immunoprecipitation. ChIP assays were performed as described previously²⁷, except that sonication of chromatin was performed with 30-s on-and-off cycles at high settings (Bioruptor, Diagenode). Details of antibodies used and PCR conditions can be found in the Supplementary Methods.

Bisulphite genomic sequencing. Bisulphite genomic sequencing was performed essentially as described²⁷. Detailed PCR amplification conditions and primer sequences are available on request.

Received 6 October; accepted 15 November 2005.

Published online 14 December 2005.

1. Egger, G., Liang, G., Aparicio, A. & Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457–463 (2004).
2. Levine, S. S., King, I. F. & Kingston, R. E. Division of labour in polycomb group repression. *Trends Biochem. Sci.* **29**, 478–485 (2004).
3. Lund, A. H. & van Lohuizen, M. Polycomb complexes and silencing mechanisms. *Curr. Opin. Cell Biol.* **16**, 239–246 (2004).
4. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
5. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (suppl.), 245–254 (2003).
6. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
7. Czermin, B. *et al.* *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
8. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
9. Muller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
10. Bachman, K. E. *et al.* Histone modifications and silencing prior to DNA methylation of a tumour suppressor gene. *Cancer Cell* **3**, 89–95 (2003).
11. Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L. & Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature Genet.* **24**, 88–91 (2000).
12. Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.* **20**, 2536–2544 (2001).
13. Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* **31**, 2305–2312 (2003).
14. Kuzmichev, A., Jenuwein, T., Tempst, P. & Reinberg, D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* **14**, 183–193 (2004).
15. Cao, R. & Zhang, Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.* **14**, 155–164 (2004).
16. Kirmizis, A. *et al.* Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* **18**, 1592–1605 (2004).
17. Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**, 277–283 (2001).
18. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).
19. Lehnertz, B. *et al.* *Suv39h*-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
20. Lewis, A. *et al.* Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nature Genet.* **36**, 1291–1295 (2004).
21. Umlauf, D. *et al.* Imprinting along the *Kcnq1* domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature Genet.* **36**, 1296–1300 (2004).
22. Rhee, I. *et al.* DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**, 552–556 (2002).
23. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell Biol.* **23**, 5594–5605 (2003).
24. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
25. Silva, J. *et al.* Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed–Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
26. Sado, T. *et al.* X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev. Biol.* **225**, 294–303 (2000).
27. Brenner, C. *et al.* Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* **24**, 336–346 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Otte for pGEX-EZH2, E. Li for DNMT3A and DNMT3B immune sera, S. Pradhan for the anti-DNMT1 antibody, and C. Beaudouin for pGEX-PLZF. We thank P. Putmans and H. N. Tran for excellent technical assistance. E.V., C.B., R.D., C.D. and L.B. were supported by the Télévie and the F.N.R.S. L.B. was also supported by the “Fondation pour la Recherche Médicale”. D.B. was supported by a Marie Curie Fellowship. F.F. is a “Chercheur Qualifié du F.N.R.S.”. This work was supported by grants from MEC-Plan Nacional de I+D+I to L.D.C., and from the “Fédération Belge contre le Cancer”, the F.N.R.S., “FB Assurances”, and “A.R.C. de la Communauté Française de Belgique” to Y.d.L. and F.F.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.F. (ffuks@ulb.ac.be).

A lever-arm rotation drives motility of the minus-end-directed kinesin Ncd

Nicholas F. Endres¹, Craig Yoshioka², Ronald A. Milligan² & Ronald D. Vale¹

Kinesins are microtubule-based motor proteins that power intracellular transport^{1,2}. Most kinesin motors, exemplified by Kinesin-1, move towards the microtubule plus end, and the structural changes that govern this directional preference have been described^{3–5}. By contrast, the nature and timing of the structural changes underlying the minus-end-directed motility of Kinesin-14 motors (such as *Drosophila* Ncd^{6,7}) are less well understood. Using cryo-electron microscopy, here we demonstrate that a coiled-coil mechanical element of microtubule-bound Ncd rotates $\sim 70^\circ$ towards the minus end upon ATP binding. Extending or shortening this coiled coil increases or decreases velocity, respectively, without affecting ATPase activity. An unusual Ncd mutant that lacks directional preference⁸ shows unstable nucleotide-dependent conformations of its coiled coil, underscoring the role of this mechanical element in motility. These results show that the force-producing conformational change in Ncd occurs on ATP binding, as in other kinesins, but involves the swing of a lever-arm mechanical element similar to that described for myosins.

Crystal structures of the Ncd dimer show a coiled coil extending directly from the pair of catalytic cores (termed the 'neck')^{9–11}, and mutagenesis experiments indicate that the proximal portion of the neck functions as the mechanical element that powers minus-end-directed motility^{9,12}. Previous cryo-electron microscopy (cryo-EM) studies of dimeric Ncd suggested that the neck extends towards the microtubule plus end in the nucleotide-free, microtubule-bound complex. This assignment was based on the position of an amino-terminal Src homology domain 3 (SH3) domain used as a marker for the neck, because the neck itself was not visible in this study¹³. Neither the neck nor the SH3 marker was observed in the presence of AMPPNP, however, which led to the proposal that ATP binding causes the Ncd neck to transition into a detached and mobile state and that the shift in the average position of this mobile neck towards the microtubule minus end is the driving force for Ncd motility¹³. A conflicting model of the nature and timing of the power stroke has been proposed on the basis of the crystal structure of a motility-deficient Ncd point mutant (NcdN600K)¹⁰, which shows the neck in a different orientation that is predicted to point towards the microtubule minus end. It has been proposed that this structure represents the nucleotide-free state of the motor and that a minus-end-directed power stroke occurs on release of ADP. Thus, the exact nature and timing of the motility-producing conformational change in Ncd remains uncertain.

Here we have used cryo-EM to investigate the position of the neck in dimeric Ncd (residues 281–700; see Methods) bound to a microtubule in the absence of nucleotide and in the presence of two mimics of an ATP-like state, AMPPNP (a non-hydrolysable ATP analogue) and ADP- AlF_4^- (a transition-state analogue). Helical image analysis of 15 protofilament microtubules was used to calculate three-

dimensional (3D) density maps (Fig. 1a–c and Methods). All three maps show two distinct globular domains of similar size and shape, which represent the two heads of the Ncd dimer. As seen before^{14–16}, only one of the two Ncd heads interacts with the microtubule. In the nucleotide-free state, an elongated density emerges from between the two heads of the Ncd dimer, and its length ($\sim 65 \text{ \AA}$) matches the expected length of the portion of the Ncd coiled coil that emerges beyond the motor domain. This same elongated density was observed in constructs with two different N-terminal tags (SH3 domain, Fig. 1; or His₆, Supplementary Fig. 1), indicating that the density was not an artefact created by a particular tag. In contrast to the previous cryo-EM study, which did not visualize the neck directly¹³ and has been subject to re-interpretation¹⁰, the clear connection of the neck density to the Ncd heads in our maps allows us to conclude unambiguously that the Ncd neck is pointing towards the plus end of the microtubule in the nucleotide-free state.

Our cryo-EM maps show a markedly different conformation of the Ncd motor in the presence of ATP analogues. With AMPPNP, the neck and unbound head of Ncd are rotated by $\sim 70^\circ$ towards the minus end of the microtubule relative to their positions in the nucleotide-free state, whereas the position of the bound head remains unchanged (Fig. 1b; see Supplementary Fig. 2 for difference maps). The neck density is slightly weak, which may reflect either the data quality or some conformational flexibility, but it can be clearly observed projecting from the two heads of the dimer. In the ADP- AlF_4^- map, the neck density is stronger and better defined (Fig. 1c). Despite these variations, an overlay of the two maps shows that the orientation of the heads and neck are identical in the AMPPNP and ADP- AlF_4^- states (Supplementary Fig. 2). Thus, our data indicate that the Ncd neck has a preferred minus-end-pointing position in the ATP-like states and is not completely random in the ATP-like states of the cycle as has been suggested¹³. These results indicate that ATP binding causes a $\sim 70^\circ$ rotation of the neck and unbound head toward the minus end of the microtubule.

We next used a docking approach to examine the relationship between the two microtubule-bound states of Ncd observed here and the two published X-rays structure of the Ncd dimer^{9,10}. We obtained an excellent fit of the wild-type Ncd/ADP crystal structure into the cryo-EM maps of the nucleotide-free state (Fig. 1d). In particular, the neck of the Ncd/ADP structure occupies the elongated density extending from the two catalytic cores (Fig. 1d), and the known microtubule-binding elements in the bound head are positioned in close proximity to tubulin^{15,17} (Supplementary Fig. 3). Keeping the bound head in the same orientation on the microtubule, we docked the crystal structure of the Ncd mutant N600K (NcdN600K)¹⁰ into the maps of the Ncd-microtubule complex in the ADP- AlF_4^- state. This structure fitted reasonably well (Fig. 1e), but was improved by a $\sim 10^\circ$ rigid-body rotation of the neck and unbound head in the

¹The Howard Hughes Medical Institute, and the Department of Cellular and Molecular Pharmacology, University of California San Francisco, 600 16th Street, San Francisco, California 94107, USA. ²Center for Integrative Molecular Biosciences, Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

NcdN600K structure (Fig. 1f). Although this slight difference suggests that the crystal structure of this non-motile Ncd mutant in solution might not be a perfect representation of the conformation of the wild-type motor bound to microtubules, our results argue that the NcdN600K structure closely approximates the microtubule-bound ATP state, rather than the nucleotide-free state as previously suggested¹⁰.

Having successfully visualized nucleotide-mediated conformational changes in wild-type Ncd, we examined an Ncd point mutant N340K (NcdN340K) that can generate motion towards either the plus end or the minus end of microtubules with roughly equal probability⁸. Because the structural basis of such bidirectional transport remains unresolved, we examined the NcdN340K–microtubule complex in different nucleotide states. In the density maps of NcdN340K in its nucleotide-free state (Fig. 1g), the shape of the unbound head is not as well defined as it is in the maps of the wild-type motor, and an elongated neck density is not observed. However, an additional disconnected density is observed at the position occupied by the N-terminal end of the neck in the wild-type motor (highlighted in Fig. 1g). One possible interpretation of this density is that the neck and unbound head in the mutant Ncd dimer occupy both the pre-stroke and post-stroke positions in the nucleotide-free state and that the 3D map reflects an average of these two positions. Consistent with this notion, averaging the wild-type nucleotide-free and ADP- AlF_4^- maps yielded a less well-defined unbound head and a disconnected neck density, similar to that seen in the NcdN340K nucleotide-free map (Fig. 1i).

In the NcdN340K/AMPPNP (Supplementary Fig. 4) and NcdN340K/ADP- AlF_4^- (Fig. 1h) maps, the neck density is completely absent, indicating an unstable neck position, and the detached head is also poorly defined (comparisons of NcdN340K/ADP- AlF_4^- and NcdN340K/AMPPNP maps show no significant differences; Supplementary Fig. 4). These data suggest a possible model of the bidirectional motion in which the mutant motor begins its ATPase cycle by binding to a microtubule with its neck oriented either towards the minus end or the plus end, and then adopts a conformationally averaged midpoint position after ATP binding. In this model, the initial direction of movement would be stochastically determined, but once motion begins in a particular direction it could continue in the same direction by virtue of cooperative effects of an ensemble of motors.

Our cryo-EM experiments suggest that Ncd uses its neck as a lever arm to generate a minus-end-directed power stroke in a manner similar to the rotation of the light-chain-binding domain in myosin II. For a lever-arm mechanism, the velocity of the motor should be proportional to the length of its lever arm¹⁸. Consistent with this prediction, a series of successive neck truncations caused a progressive decrease in velocity in a microtubule-gliding assay, but did not affect enzymatic turnover (ATPase catalytic rate constant, k_{cat} ; Fig. 2a). This finding is consistent with previous work on Ncd truncations^{10,19}, although ATPase activity was not examined in those studies. Truncation experiments are difficult to interpret, however, because the loss of protein structure could damage motor function in unanticipated ways. We therefore sought to increase

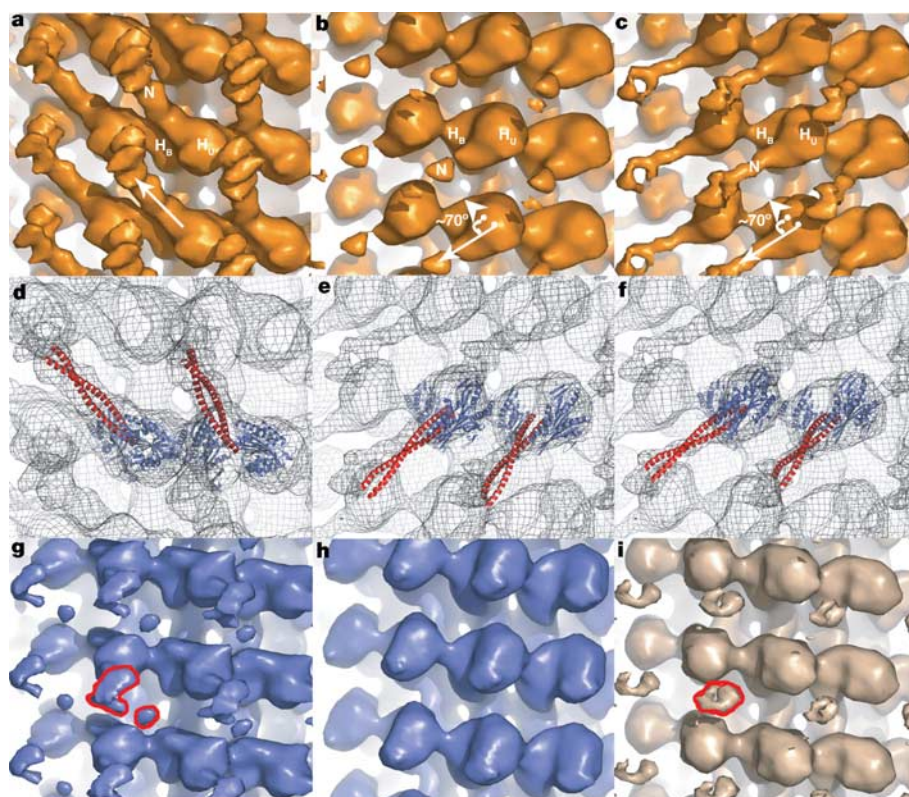


Figure 1 | 3D maps of Ncd-microtubule complexes by cryo-EM. **a–c**, Wild-type Ncd in the absence of nucleotide (**a**) and in the presence of the nucleotide analogues Mg^{2+} -AMPPNP (5 mM, **b**) and Mg^{2+} -ADP- AlF_4^- (5 mM, **c**). Putative density for the bound head, unbound head and neck region are marked as H_b , H_u and N , respectively. Arrows indicate the rotation of the unbound head and neck between the nucleotide-free and ATP-like states. **d–f**, Ncd crystal structures docked into wild-type maps. Shown is the Ncd-ADP structure⁹ docked into the nucleotide-free maps (**d**). The NcdN600K structure¹⁰ (**e**) and the NcdN600K structure with a 10° rotation of the unbound head and neck about Gly 347 (the residue marking

the boundary between the neck and the catalytic core) (**f**), are shown docked to the Mg^{2+} -ADP- AlF_4^- maps. **g, h**, Cryo-EM maps for the bidirectional mutant NcdN340K, without nucleotide (**g**) and with 5 mM Mg^{2+} -ADP- AlF_4^- (**h**). 5 mM Mg^{2+} -AMPPNP produces an equivalent map to 5 mM Mg^{2+} -ADP- AlF_4^- (Supplementary Fig. 3). **i**, Density map generated by averaging the Ncd wild-type nucleotide-free and Mg^{2+} -ADP- AlF_4^- data. Detached density in NcdN340K nucleotide-free map (**g**) and in Ncd average map (**i**) circled in red. All figures are oriented so that the plus end of the microtubule axis is at the top of the page. Figs were generated with Pymol (Delano Scientific).

velocity by extending the length of the neck with a four-heptad leucine zipper coiled-coil motif ('LZ extension'). As expected for a lever-arm model, fusion of this LZ extension to the native Ncd neck at three different positions increased microtubule gliding velocity without changing ATPase k_{cat} (Fig. 2a). This increase in velocity was not observed when a flexible glycine-serine linker was inserted between the LZ extension and the native Ncd neck (Fig. 2a), suggesting that the LZ extension increases Ncd velocity by extending the length of the mechanical element and not by some other mechanism. The compiled velocity data from the seven truncated or extended neck constructs show that microtubule gliding velocity is proportional to the predicted length of the neck, regardless of whether native or nonnative (LZ extension) residues were used, but ATPase k_{cat} remains unaffected (Fig. 2b, c). Taken together, these data support the notion that a lever-arm rotation of the Ncd neck powers minus-end-directed motility.

A model of Ncd motility invoking the rotation of the neck suggests that the unbound head may not be necessary to generate motility. To test this notion, we prepared a single-headed Ncd heterodimer (N280_Het; Fig. 2a) in which one polypeptide consisted of an intact Ncd catalytic core and neck (residues 280–700) and the second polypeptide consisted of the neck region alone (residues 281–347;

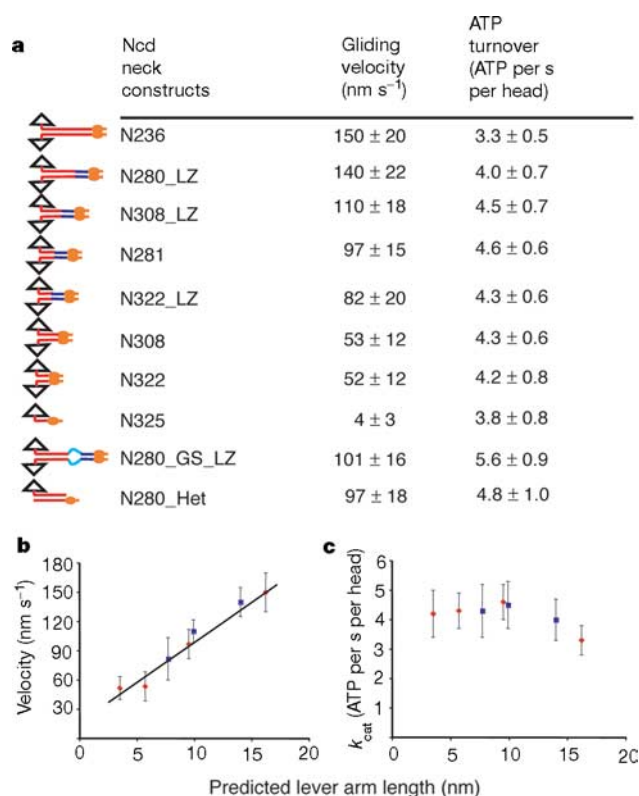


Figure 2 | Ncd mutants with truncated or extended necks. **a**, Gliding assay velocities and ATPase k_{cat} (in ATP molecules per head per s). Velocity data are the mean ± s.d. ($n > 150$); ATPase k_{cat} data are the weighted average and errors obtained from the fits of two independent ATPase experiments from at least two protein preparations. Numbers in the construct name represent the starting residue for the native Ncd residues in the construct. N-terminal LZ extensions (29 residues) and the flexible Gly-Ser linker (12 residues) are labelled LZ and GS, respectively (see Methods). The domain structures of the constructs are shown on the left with motor cores in black, native neck residues in red, LZ extensions in dark blue, the flexible linker in light blue, and the biotin tag in orange. **b**, **c**, Velocities of gliding movement (mean ± s.d.; **b**) and ATP k_{cat} (weighted average and errors; **c**) plotted against predicted neck length of each construct. Data points representing constructs containing only native neck residues are red, and data from constructs with LZ extensions are blue.

Supplementary Fig. 5a and Methods). This motor elicited microtubule gliding at a velocity comparable to that of the normal two-headed Ncd homodimer with a similar ATPase k_{cat} (Fig. 2a). Thus, although our cryo-EM data show that the unbound head rotates along with the neck, the functional data from the heterodimer indicate that contacts between the neck and unbound head are not essential for the mechanism and that the neck alone is sufficient to act as a lever arm. Studies have also shown that a naturally occurring Kinesin-14 heterodimer in yeast (the Kar3p-Cik1p complex, a motor polypeptide in complex with a motor-less coiled coil²⁰) is an active, force-producing motor^{21,22}.

To determine whether the lever-arm motion of the Ncd neck requires a stable coiled-coil interaction, we also tested the motility of an Ncd monomer construct (N325). This construct showed >25-fold reduced motility compared with the single-headed heterodimer, but had an ATPase activity similar to that of the other constructs (Fig. 2a and Supplementary Fig. 5). Thus, a stable coiled coil is required for optimal function of the motor, as would be expected for a lever-arm model.

On the basis of our structural and functional data, we propose the following model of Ncd motility (Fig. 3). Ncd from solution binds to microtubules using one of its heads, triggering ADP release²³. The excellent fit of the Ncd/ADP structure to the nucleotide-free maps suggests that microtubule binding and ADP release do not produce large-scale conformational changes in the Ncd dimer. Our cryo-EM data suggest that ATP binding leads to a ~70° rigid-body rotation of the neck that produces a minus-end-directed displacement. A subsequent protein isomerization step, possibly before phosphate release, triggers the formation of a weakly bound state and the dissociation of Ncd from the microtubule^{24,25}. The neck lever arm can then return to its pre-power stroke position after dissociating from the microtubule, thereby completing the cycle (Fig. 3). Although this overall scheme is supported by our data, questions remain open about the proposed lever-arm mechanism. Specifically, although our data unequivocally show a preferred position of the lever arm in the AMPPNP and ADP-AlF₄⁻ states, the weaker density in our AMPPNP maps suggests that this post-powerstroke state

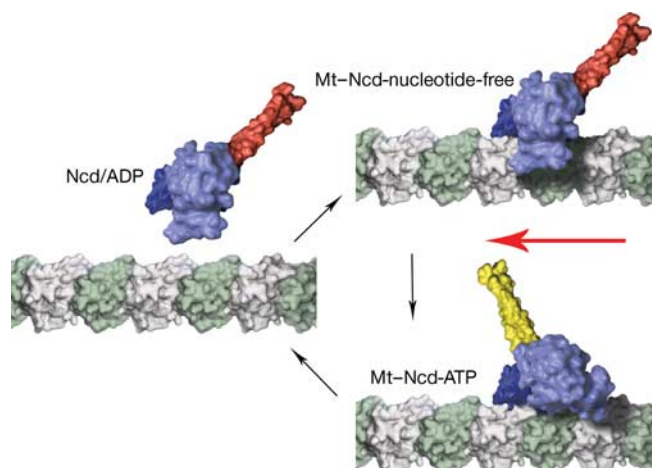


Figure 3 | Model of the Ncd motility cycle. The model is based on surface representations of the docked Ncd structures (Fig. 1d, f). The microtubule is oriented so that the plus end is on the right. In this model, ATP binding causes a rotation of the neck (coloured red in the nucleotide-free and yellow in the ATP state) that leads to a minus-end displacement along the microtubule (indicated by the red arrow). After this lever-arm rotation, the motor releases from the microtubule after nucleotide hydrolysis but probably before phosphate release^{24,25}. The released motor then returns to its pre-powerstroke position so that the cycle can repeat. Images rendered from atomic structures by Graham Johnson (fiVth media: <http://www.FiVth.com>).

may not be completely rigid and fixed in position, as envisaged by classical swinging crossbridge models of myosin. Future work on this issue will require dynamic measurement of the lever-arm position in different nucleotide states with high spatial and temporal resolution.

Our work shows that the mechanical event in the minus-end-directed Ncd (rotation of the coiled-coil neck) is coupled to the same step of the ATPase cycle (ATP binding) as the mechanical event in the plus-end-directed kinesins (neck linker docking). Thus, reversal of direction in Kinesin-14 motors is accomplished by the evolution of a unique mechanical element that can take advantage of existing conformational changes in the catalytic core, as is also true for direction reversal by the myosin VI motor²⁶. Unlike conventional kinesin, which is built for long-distance processive movement, Ncd is a nonprocessive motor^{27,28} designed for microtubule crossbridging and tension development in meiotic or mitotic spindles²⁹. In this regard, the functions of Ncd are more similar to the tension-generating myosin II motors in muscle. Thus, Ncd and muscle myosin convergently evolved a similar strategy for motility involving a large-scale rotation of an elongated lever and the primary use of only one of the two heads in the motor dimer.

METHODS

Cloning and protein preparation. The constructs used for motility and ATPase assays had an N-terminal pET104 BioEase tag (Invitrogen) for biotinylation and the cryo-EM constructs had an N-terminal SH3 domain cloned from human Nck1 fused to residue 281 of Ncd. All constructs had an N-terminal His₆ tag for purification. The LZ extensions (Fig. 2) contained the yeast GCN4 sequence (VKQLEDKVEELLSKNYHLENEVARLKKIV), and N280_LZ_GS contained a glycine-serine linker (GGGSGGGSGGG). N280_Het was prepared by coexpressing a biotin- and His₆-tagged neck domain (281–347) with an untagged motor domain (280–700) on the same pET-17b plasmid (Invitrogen) using a single T7 promoter. Proteins were expressed and purified from a Ni²⁺-NTA agarose column (Qiagen) as described³⁰. For motility and ATPase assays, a microtubule affinity purification step, similar to that reported previously³⁰, was used to select for active motors. N280_Het required an additional gel filtration purification step (Supplementary Fig. 5).

Cryo-EM and helical image analysis. Ncd (4–7 mg ml⁻¹) was dialysed against 25 mM MOPS (pH 7.25), 100 mM NaCl, 2 mM MgCl₂, 1 mM EGTA and 1 mM dithiothreitol and was centrifuged (100,000g, 15 min) to remove any precipitate. Frozen grids containing microtubules (5 mg ml⁻¹) and Ncd were prepared as described¹⁵ with a final concentration of nucleotide of 5 mM or 1 U of apyrase. Imaging and image analysis were carried out essentially as described¹⁵, with an FEI CM200FEG electron microscope and Gatan cold stage. The number of data sets averaged and the total number of asymmetric units contributing to each 3D map were as follows: Ncd/AMPPNP, 20 data sets, 17,600 particles; Ncd/nucleotide-free, 24 data sets, 17,100 particles; Ncd/ADP-ATP, 26 data sets, 23,700 particles; Ncd(N340K)/AMPPNP, 20 data sets, 12,150 particles; Ncd(N340K)/ATP, 20 data sets, 13,000 particles; Ncd(N340K)/nucleotide-free, 29 data sets, 17,250 particles. For the figures, all 3D data sets were fitted and scaled to a reference created by averaging the microtubule–Ncd/ATP and microtubule–Ncd/nucleotide-free data (Fig. 1i).

Microtubule gliding and ATPase assays. For gliding assays, glass slides were treated with 0.5 mg ml⁻¹ of biotinylated bovine serum albumin (Pierce) and then 0.5 mg ml⁻¹ of streptavidin (Pierce) before the motor (200–400 nM for homodimers, 2 μ M for heterodimer) was added. Gliding velocities of rhodamine-labelled microtubules were measured as described³⁰ in motor buffer (25 mM MOPS, 100 mM NaCl, 1 mM EGTA, 2 mM MgCl₂ and 5% sucrose; pH 7). ATP hydrolysis rates were measured by an Enzcheck assay kit (Molecular Probes) using 10–40 nm of motor and 0–50 μ M microtubules in motor buffer with 10 μ M paclitaxel and NaCl reduced to 25 mM. Hydrolysis of ATP was plotted against microtubule concentrations, and data were fitted to a Michaelis–Menton equation to determine k_{cat} .

Received 2 September; accepted 23 September 2005.

Published online 28 December 2005.

- Hirokawa, N. & Takemura, R. Kinesin superfamily proteins and their various functions and dynamics. *Exp. Cell Res.* **301**, 50–59 (2004).
- Sharp, D. J., Rogers, G. C. & Scholey, J. M. Microtubule motors in mitosis. *Nature* **407**, 41–47 (2000).

- Rice, S. *et al.* A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
- Asbury, C. L., Fehr, A. N. & Block, S. M. Kinesin moves by an asymmetric hand-over-hand mechanism. *Science* **302**, 2130–2134 (2003).
- Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin walks hand-over-hand. *Science* **303**, 676–678 (2004).
- McDonald, H. B., Stewart, R. J. & Goldstein, L. S. The kinesin-like ncd protein of *Drosophila* is a minus end-directed microtubule motor. *Cell* **63**, 1159–1165 (1990).
- Endow, S. A., Henikoff, S. & Soler-Niedziela, L. Mediation of meiotic and early mitotic chromosome segregation in *Drosophila* by a protein related to kinesin. *Nature* **345**, 81–83 (1990).
- Endow, S. A. & Higuchi, H. A mutant of the motor protein kinesin that moves in both directions on microtubules. *Nature* **406**, 913–916 (2000).
- Sablin, E. P. *et al.* Direction determination in the minus-end-directed kinesin motor ncd. *Nature* **395**, 813–816 (1998).
- Yun, M. *et al.* Rotation of the stalk/neck and one head in a new crystal structure of the kinesin motor protein, Ncd. *EMBO J.* **22**, 5382–5389 (2003).
- Kozielewski, F., De Bonis, S., Burmeister, W. P., Cohen-Addad, C. & Wade, R. H. The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations. *Struct. Fold. Des.* **7**, 1407–1416 (1999).
- Endow, S. A. & Waligora, K. W. Determinants of kinesin motor polarity. *Science* **281**, 1200–1202 (1998).
- Wendt, T. G. *et al.* Microscopic evidence for a minus-end-directed power stroke in the kinesin motor ncd. *EMBO J.* **21**, 5969–5978 (2002).
- Hirose, K., Lockhart, A., Cross, R. A. & Amos, L. A. Three-dimensional cryoelectron microscopy of dimeric kinesin and ncd motor domains on microtubules. *Proc. Natl Acad. Sci. USA* **93**, 9539–9544 (1996).
- Sosa, H. *et al.* A model for the microtubule–Ncd motor protein complex obtained by cryo-electron microscopy and image analysis. *Cell* **90**, 217–224 (1997).
- Arnall, I., Metz, F., DeBonis, S. & Wade, R. H. Three-dimensional structure of functional motor proteins on microtubules. *Curr. Biol.* **6**, 1265–1270 (1996).
- Woehlke, G. *et al.* Microtubule interaction site of the kinesin motor. *Cell* **90**, 207–216 (1997).
- Uyeda, T. Q., Abramson, P. D. & Spudis, J. A. The neck region of the myosin motor domain acts as a lever arm to generate movement. *Proc. Natl Acad. Sci. USA* **93**, 4459–4464 (1996).
- Stewart, R. J., Thaler, J. P. & Goldstein, L. S. Direction of microtubule movement is an intrinsic property of the motor domains of kinesin heavy chain and *Drosophila* ncd protein. *Proc. Natl Acad. Sci. USA* **90**, 5209–5213 (1993).
- Barrett, J. G., Manning, B. D. & Snyder, M. The Kar3p kinesin-related protein forms a novel heterodimeric structure with its associated protein Cik1p. *Mol. Biol. Cell* **11**, 2373–2385 (2000).
- Sproul, L. R., Anderson, D. J., Mackey, A. T., Saunders, W. S. & Gilbert, S. P. Cik1 targets the minus-end kinesin depolymerase kar3 to microtubule plus ends. *Curr. Biol.* **15**, 1420–1427 (2005).
- Chu, H. M. *et al.* Kar3 interaction with Cik1 alters motor structure and function. *EMBO J.* **24**, 3214–3223 (2005).
- Lockhart, A. & Cross, R. A. Origins of reversed directionality in the ncd molecular motor. *EMBO J.* **13**, 751–757 (1994).
- Pechatnikova, E. & Taylor, E. W. Kinetic mechanism of monomeric non-claret disjunctional protein (Ncd) ATPase. *J. Biol. Chem.* **272**, 30735–30740 (1997).
- Foster, K. A., Correia, J. J. & Gilbert, S. P. Equilibrium binding studies of non-claret disjunctional protein (Ncd) reveal cooperative interactions between the motor domains. *J. Biol. Chem.* **273**, 35307–35318 (1998).
- Menetrey, J. *et al.* The structure of the myosin VI motor reveals the mechanism of directionality reversal. *Nature* **435**, 779–785 (2005).
- deCastro, M. J., Fondecave, R. M., Clarke, L. A., Schmidt, C. F. & Stewart, R. J. Working strokes by single molecules of the kinesin-related microtubule motor ncd. *Nature Cell Biol.* **2**, 724–729 (2000).
- Pechatnikova, E. & Taylor, E. W. Kinetics processivity and the direction of motion of Ncd. *Biophys. J.* **77**, 1003–1016 (1999).
- Sharp, D. J., Yu, K. R., Sisson, J. C., Sullivan, W. & Scholey, J. M. Antagonistic microtubule-sliding motors position mitotic centrosomes in *Drosophila* early embryos. *Nature Cell Biol.* **1**, 51–54 (1999).
- Case, R. B., Pierce, D. W., Hom-Booher, N., Hart, C. L. & Vale, R. D. The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain. *Cell* **90**, 959–966 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Rice, J. Kardon and J. Epstein for initial work on this project. This work was supported by NIH grants (to R.D.V. and R.A.M.). Some of the work presented here was conducted at the National Resource for Automated Molecular Microscopy, which is supported by the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.D.V. (vale@cmp.ucsf.edu).

Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB

Bomina Yu¹, William C. Edstrom¹, Jordi Benach¹, Yoshitomo Hamuro², Patricia C. Weber², Brian R. Gibney³ & John F. Hunt¹

Nucleic acid damage by environmental and endogenous alkylation reagents creates lesions that are both mutagenic and cytotoxic, with the latter effect accounting for their widespread use in clinical cancer chemotherapy^{1,2}. *Escherichia coli* AlkB^{3–9} and the homologous human proteins ABH2 and ABH3 (refs 5, 7) promiscuously repair DNA and RNA bases damaged by S_N2 alkylation reagents, which attach hydrocarbons to endocyclic ring nitrogen atoms (N1 of adenine and guanine and N3 of thymine and cytosine)^{3,4,10–15}. Although the role of AlkB in DNA repair has long been established based on phenotypic studies, its exact biochemical activity was only elucidated recently after sequence profile analysis revealed it to be a member of the Fe-oxoglutarate-dependent dioxygenase superfamily. These enzymes use an Fe(II) cofactor and 2-oxoglutarate co-substrate to oxidize organic substrates. AlkB hydroxylates an alkylated nucleotide base to produce an unstable product that releases an aldehyde to regenerate the unmodified base. Here we have determined crystal structures of substrate and product complexes of *E. coli* AlkB at resolutions from 1.8 to 2.3 Å. Whereas the Fe-2-oxoglutarate dioxygenase core matches that in other superfamily members, a unique subdomain holds a methylated trinucleotide substrate into the active site through contacts to the polynucleotide backbone. Amide hydrogen exchange studies and crystallographic analyses suggest that this substrate-binding ‘lid’ is conformationally flexible, which may enable docking of diverse alkylated nucleotide substrates in optimal catalytic geometry. Different crystal structures show open and closed states of a tunnel putatively gating O₂ diffusion into the active site. Exposing crystals of the anaerobic Michaelis complex to air yields slow but substantial oxidation of 2-oxoglutarate that is inefficiently coupled to nucleotide oxidation. These observations suggest that protein dynamics modulate redox chemistry and that a hypothesized migration of the reactive oxy-ferryl ligand on the catalytic Fe ion may be impeded when the protein is constrained in the crystal lattice.

The likely catalytic mechanism of AlkB is shown as a schematic diagram in Fig. S1 in the Supplementary Information, which also describes the engineering of a crystallizable and enzymatically active 11-residue, amino-terminal truncation of *E. coli* AlkB (AlkB-ΔN11). The crystal structure of its anaerobic Michaelis complex with Fe(II), 2-oxoglutarate (2OG) and the substrate dT-(1-me-dA)-dT¹⁴ was solved using multiwavelength anomalous diffraction experiments on the Fe cofactor (Fig. 1a). The carboxy-terminal Fe-2OG dioxygenase core comprises a double-stranded β-helix in which eight β-strands form a ‘jellyroll’ fold closely matching that observed in other superfamily members^{16–18} (Fig. 1b) (the clavamate synthase superfamily in SCOP¹⁹). This core binds Fe and 2OG using residues that are invariant in the superfamily (His 131, Asp 133 and His 187

ligating Fe and Arg 204 and Arg 210 forming salt bridges to the carboxylates of 2OG) (Fig. 1a; see also Supplementary Fig. S2A).

The 90 N-terminal residues from *E. coli* AlkB form structural elements that are not shared with other Fe-2OG dioxygenases (Fig. 1b). These AlkB-specific elements form a β-strand and α-helix that pack onto the surface of the dioxygenase core followed by a β-sheet-forming lid subdomain that contacts the trinucleotide substrate and covers the active site. Except for contacts involving the Fe-2OG-ligating residues and their immediate neighbours in the dioxygenase core, most of the contacts to the dT-(1-me-dA)-dT substrate come from this subdomain, which we therefore designate the ‘nucleotide-recognition lid’ (Fig. 2; see also Supplementary Fig. 2A).

Convergent evidence indicates that the nucleotide-recognition lid is flexible but that its flexibility is reduced upon binding the alkylated trinucleotide substrate (Fig. 2b; see also Supplementary Figs S2 and S4A, B). High-resolution backbone amide ¹H/²H (H/D) exchange studies conducted in the absence of trinucleotide substrate (Supplementary Fig. S2A) indicate that this region undergoes relatively rapid exchange and is therefore dynamically flexible. However, its H/D exchange rates and flexibility are reduced upon binding dT-(1-me-dA)-dT. In a nucleotide-free crystal structure of AlkB-ΔN11 (Supplementary Table S2 and Supplementary Figs S2B and S4B), the protein segments flanking the nucleotide-binding site have elevated B-factors and enhanced mobility compared to the protein core. These segments change backbone conformation (Fig. 2b; see also Supplementary Fig. S4A) and the loops in the nucleotide-recognition lid exhibit reduced B-factors (Supplementary Fig. S2B) in the crystal structure of the anaerobic Michaelis complex with dT-(1-me-dA)-dT, consistent with reduced mobility upon binding nucleotide substrate. Arguments elaborated below suggest that AlkB has developed an elegant molecular design in which these flexible protein segments recognize invariant nucleic acid backbone features to produce high-affinity docking of chemically diverse modified bases into its conserved catalytic core.

The nucleotide backbone is bound in a cooperative hydrogen-bonding (H-bonding) network in which the two phosphates 5′ to the alkylated adenine base interact with Thr 51, Tyr 76 and Arg 161 (Fig. 2b), which are invariant in eubacterial AlKBs (Supplementary Fig. S2A). The first two residues are located in flexible segments of the nucleotide-recognition lid, whereas Arg 161 comes from a flexible surface loop comprising residues 159–164 in the dioxygenase core (Fig. 2b; see also Supplementary Figs S2 and S4A, B). The centre of the H-bonding network is Tyr 76, which forms a bidentate H-bond bridging the two phosphates. Arg 161 makes a salt bridge to the 5′ phosphate of the alkylated adenosine (Fig. 2a, b), explaining the importance of this phosphate for substrate affinity¹⁴. However, there are no other salt bridges to the trinucleotide, and the substrate

¹Department of Biological Sciences and Northeast Structural Genomics Consortium, 702A Fairchild Center, MC2434, Columbia University, New York, New York 10027, USA.

²ExSAR Corporation, 11 Deer Park Drive, Suite 103, Monmouth Junction, New Jersey 08852, USA. ³Department of Chemistry, 3000 Broadway, MC 3121, Columbia University, New York, New York 10027, USA.

binding pocket does not carry significant net charge (Supplementary Fig. S4C). The base stacking interaction between the terminal bases in the trinucleotide (Fig. 2b) may stabilize the backbone conformation recognized by AlkB in which the central alkylated base is flipped out relative to the flanking bases²⁰, which could contribute to the bacterial enzyme's greater activity on single-stranded substrates^{5,7}.

The alkylated 1-methyladenine base is bound in a deep, predominantly hydrophobic cavity, where it is sandwiched between invariant residues Trp 69 from the nucleotide-recognition lid and His 131, an Fe-ligating residue in the dioxygenase core (Fig. 2). Whereas the aromatic stacking interaction with Trp 69 can be made with any substrate base, an H-bond between conserved residue Asp 135 and the N6 group of adenine can also be made with cytidine but not guanine or thymine, explaining AlkBs higher efficiency in demethylating the first two^{10–12}. There is little unoccupied volume in the binding cavity (Fig. 2c), and the target 1-methyl group directly

contacts its molecular surface as well as an oxygen atom on the 1-carboxylate of 2OG (Fig. 3a).

A small packing void adjacent to the 1-methyl group may enable 1-N6-ethenoadenine¹⁵ to be bound and repaired in equivalent geometry. However, the repair of bulkier alkyl modifications or other modified bases, especially the smaller pyrimidines, is likely to require reorientation of the base and movement of the attached polynucleotide backbone to maintain the precise alignment of the target alkyl atom required for efficient oxidation. The flexibility of the protein segments holding the nucleic acid substrate into the active site (Fig. 2b; see also Supplementary Fig. S4A, B) should enable any requisite adjustments to occur while conserving the molecular interactions with invariant features on the polynucleotide backbone (Fig. 2a, b).

The molecular mechanism of Fe-2OG dioxygenases has been investigated extensively^{21–24}. Cleavage of molecular oxygen bound

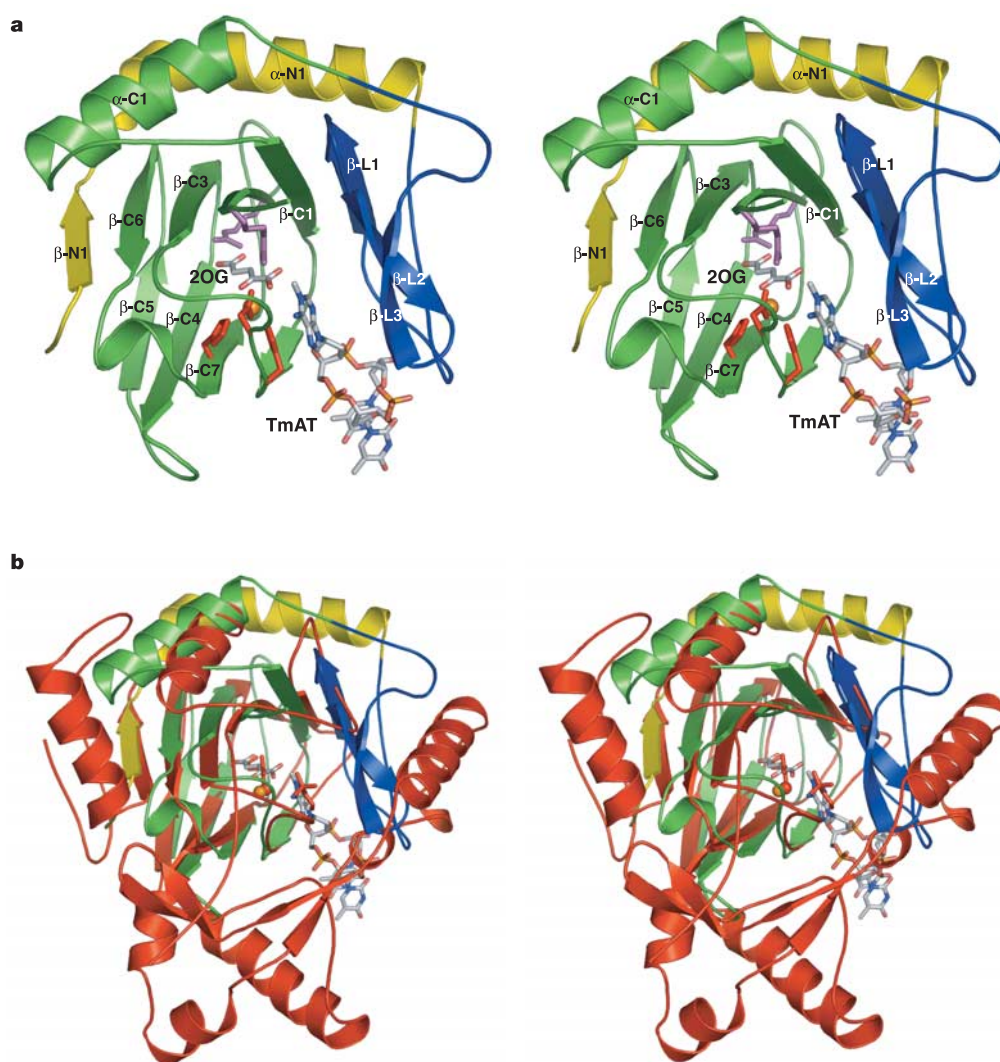


Figure 1 | Crystal structure of the anaerobic Michaelis complex of *E. coli* AlkB- Δ N11 with Fe(II), 2OG and a methylated trinucleotide. **a, Stereo ribbon diagram with the backbone coloured and the 2° structural elements labelled according to subdomain organization (with the N-terminal extension (N) in yellow, nucleotide-recognition lid (L) in blue, and catalytic core (C) in green as in the sequence-structure alignment in Supplementary Fig. S2A). Most of the blue segment is protected against amide H/D exchange by dT-(1-me-dA)-dT substrate binding (Supplementary Fig. S2A). The sphere representing the Fe cofactor is coloured orange, whereas atoms in 2OG and dT-(1-me-dA)-dT are coloured according to atomic identity**

(carbon, white; oxygen, red; nitrogen, blue; and phosphorous, orange). Invariant side chains in Fe-2OG dioxygenases are coloured red or magenta depending on whether they interact with Fe or 2OG, respectively. **b**, Least-squares superposition of 80 out of 211 C α atoms in AlkB with a root mean square deviation of 2.1 Å with the equivalent atoms in the taurine oxidase TauD¹⁶ (Protein Data Bank 1OS7; protein backbone and ligands coloured red). In TauD, the 1-carboxylate of 2OG interacts with Fe in the alternative geometry observed in crystal structures of some Fe-2OG dioxygenases before O₂-analogue binding¹⁸.

to Fe is coupled to oxidation of 2OG to yield succinate and CO₂ (Supplementary Fig. S1). This reaction is postulated to generate a reactive oxy-ferryl intermediate (Fe(IV) = O in Supplementary Fig. S1) that hydroxylates the alkyl moiety on the nucleotide

base^{25–27}. The resulting carbinoliminium linkage to the base is unstable and spontaneously releases an aldehyde product (formaldehyde for a methylated substrate) to regenerate the unmodified base.

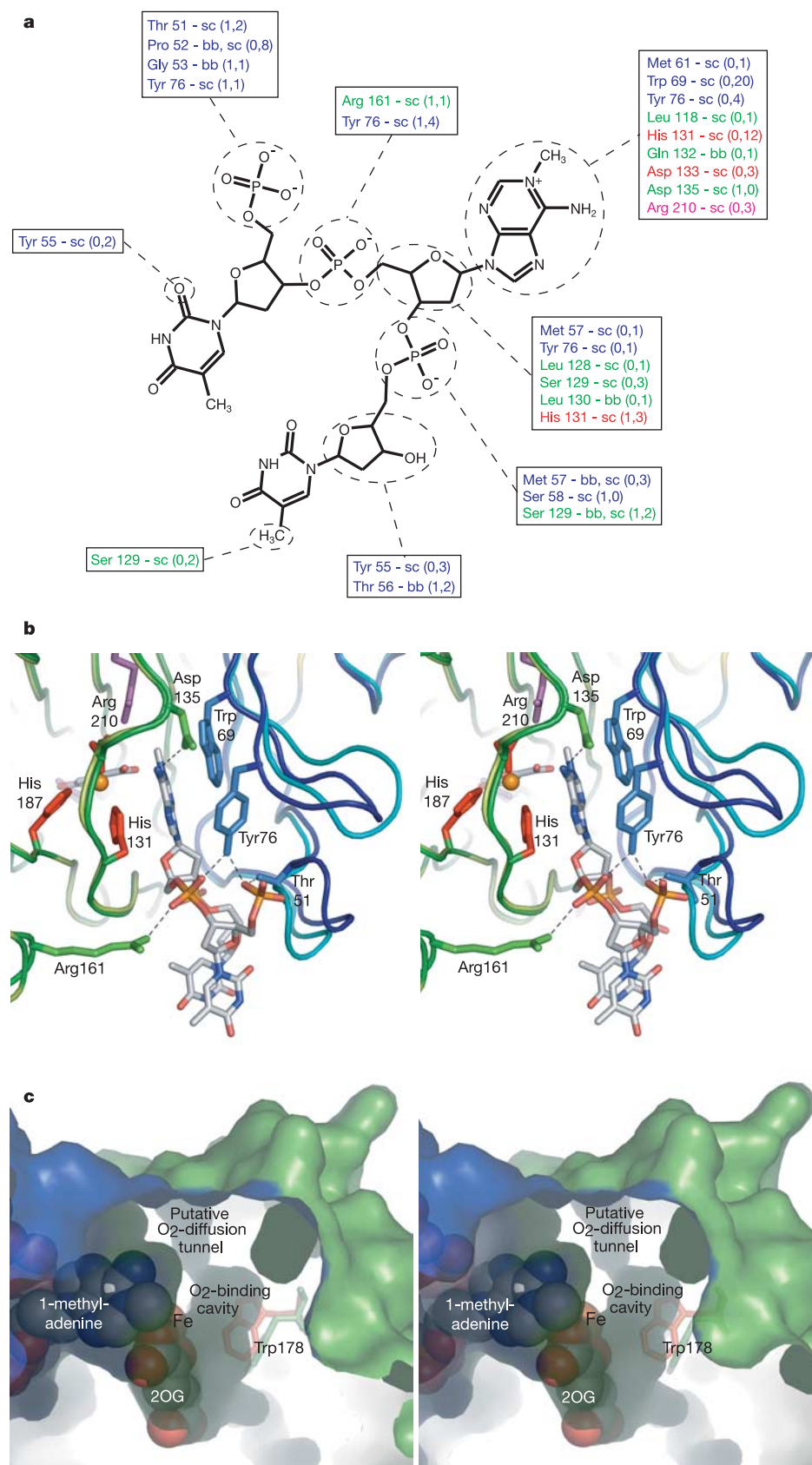


Figure 2 | Nucleotide substrate recognition and active site cavities in the anaerobic Michaelis complex. **a**, Contact diagram showing protein backbone (bb) and side chain (sc) interactions with dT-(1-me-dA)-dT, with the number of H-bonds (≤ 3.5 Å) indicated then the number of additional van der Waals contacts (≤ 3.9 Å) indicated in parentheses. **b**, Stereo pair showing the protein backbone from both the anaerobic Michaelis complex in space group *P*₄₃ (darker colours) and the nucleotide-free complex with Fe and succinate in space group *P*₁ (lighter colours). Side chains that are not invariant in Fe-2OG dioxygenases are coloured according to the subdomain of origin (as in Fig. 1). **c**, Stereo pair showing cavities near the active site in the structure of the anaerobic Michaelis complex after 2 h of exposure to air. The molecular surface of the protein is coloured according to the subdomain of origin. Trp 178 is shown in the different conformations observed in the anaerobic Michaelis complex (green) and the Co(II)-2OG-dT-(1-me-dA)-dT complex (red). For additional information, see the legend to Supplementary Fig. S6, which shows an equivalent view of the Co(II) complex in which the putative O₂-diffusion tunnel is closed.

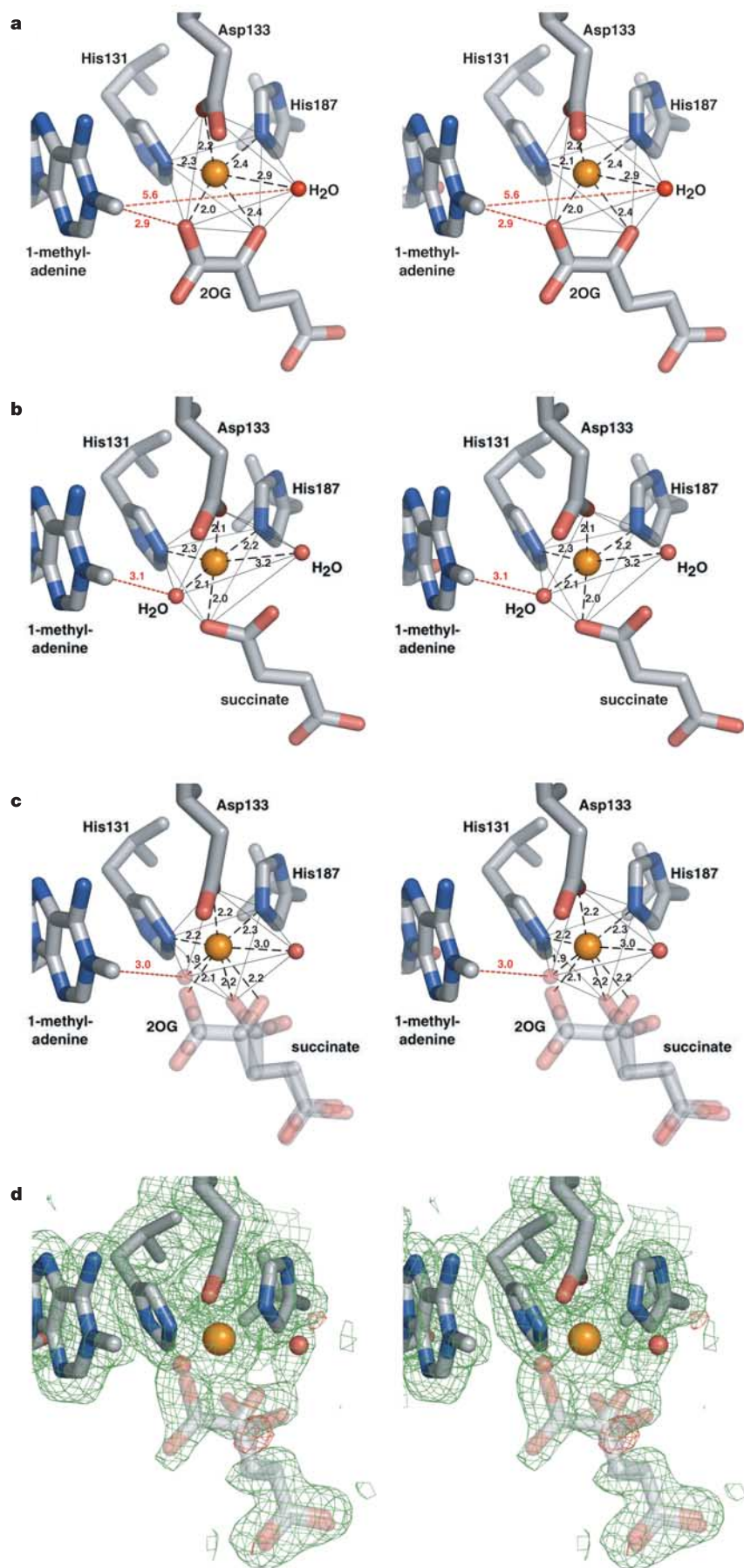


Figure 3 | Stereo pairs showing active site stereochemistry in alternative ligand complexes of AlkB- Δ N11. Ligands and side chains are coloured according to atomic identity as in Fig. 1. **a**, The anaerobic Michaelis complex has all of the octahedral coordination sites on the Fe cofactor occupied by protein or 2OG atoms except for a single site occupied by a crystallographic water, which must be replaced by O₂ to initiate oxidation (Supplementary Fig. S1). **b**, Equivalent view of the structure co-crystallized with Fe, succinate and dT-(1-me-dA)-dT (Supplementary Table S2). **c**, Equivalent view of the structure in which the anaerobic Michaelis complex was exposed to oxygen for 2 h. Unbiased electron density maps (Supplementary Fig. S5B) and occupancy refinement (Supplementary Table S3) both support the conclusion that most of the 2OG has been oxidized to succinate but that the adenine base remains largely methylated after short-term *in situ* oxidation. The alternative ligands are shown in semi-transparent rendering with the degree of transparency scaled according to refined occupancy. **d**, Refined $2F_o - F_c$ (green) and $F_o - F_c$ (red) electron density maps for the structure shown in **c** contoured at +1 σ and -3 σ , respectively. There are no $F_o - F_c$ peaks $\geq +3\sigma$ in this region.

In the anaerobic Michaelis complex of AlkB, the primary coordination sphere of the non-haem iron contains a single water molecule, the 2OG co-substrate bound in a bidentate fashion via the 2-keto and 1-carboxylate oxygens, and a facial interaction with the side chains of His 131, Asp 133 and His 187, all invariant residues in Fe-2OG dioxygenases²³ (Fig. 3a; see also Supplementary Fig. S1). Equivalent octahedral coordination geometry is observed in oxygen-resistant AlkB substrate complexes with Fe(II) replaced by Co(II) or Mn(II) (Supplementary Table S2 and data not shown).

To initiate oxidation, O₂ is believed to displace the water molecule from the primary coordination sphere of Fe²³. This site in AlkB is stereochemically equivalent to the established site of O₂-analogue (NO) binding in clavamine synthase¹⁸ (CAS), another Fe-2OG dioxygenase, and is ideally positioned for oxidation of 2OG. However, completion of the first phase of the reaction in this geometry will generate the reactive oxy-ferryl ligand at the coordination site initially occupied by the displaced water, which is 5.6 Å away from the target methyl carbon on dT-(1-me-dA)-dT (Fig. 3a), so the oxy-ferryl ligand will need to migrate to the adjacent coordination position, which is 2.9 Å from the target methyl carbon (Fig. 3a), in order to complete the second phase of the reaction. Approximately half of the Fe-2OG dioxygenases of known structure show AlkB-like octahedral Fe coordination in the anaerobic Michaelis complex^{24,28}. The remainder show an alternative conformation of 2OG that exchanges the position of its 1-carboxylate oxygen with that of the bound water (like TauD¹⁶ in Fig. 1b), which would eliminate the need for oxy-ferryl ligand migration if O₂ were to bind to the site occupied by water in these structures. However, CAS also shows this alternative 2OG conformation in the anaerobic Michaelis complex, but the conformation of its bound 2OG changes upon binding the O₂ analogue NO to yield an inhibited catalytic complex with the same Fe coordination geometry as AlkB¹⁸. Therefore, oxy-ferryl ligand migration may be a general feature of Fe-2OG dioxygenase reaction chemistry.

Migration is not possible while 2OG is bound because an oxygen atom from its 1-carboxylate occupies the destination coordination site. However, co-crystallization of AlkB with succinate, the product of 2OG oxidation, yields a structure with monodentate coordination of Fe by the product plus a second water molecule bound at the site previously occupied by the 1-carboxylate of 2OG (Fig. 3b; see also Supplementary Table S2). Therefore, oxy-ferryl migration into this coordination site adjacent to the target methyl group can occur after release of CO₂ from 2OG¹⁸.

Studies of the oxidation reaction *in situ* in AlkB-ΔN11 crystals suggest that protein dynamics may have a role in modulating oxy-ferryl migration as well as O₂ diffusion into the active site (Fig. 3; see also Supplementary Figs S5 and S6). The truncated enzyme turns over dT-(1-me-dA)-dT at a rate of 4.5 per minute in solution (Supplementary Fig. S3), similar to the full-length enzyme¹⁴. However, it catalyses oxidation slowly and inefficiently when constrained in the lattice (Fig. 3c, d; see also Supplementary Table S3 and Supplementary Fig. S5B, C). Anaerobically grown crystals exposed to air for 9 days show ~90% conversion of 2OG to succinate but only ~40% demethylation of 1-methyladenine based on occupancy refinement (Supplementary Table S3) and analysis of unbiased electron density (Supplementary Fig. S5C). Equivalent analyses of crystals exposed to air for 2 h show that although ~60% of the 2OG is oxidized, little 1-methyladenine is demethylated (Fig. 3c, d; see also Supplementary Table S3 and Supplementary Fig. S5B). Therefore, oxidation proceeds slowly in the crystals. Furthermore, because dT-(1-me-dA)-dT exchange is unlikely in the lattice, especially in the absence of succinate/2OG exchange, most turnovers can be inferred not to productively complete the second phase of the reaction. This inefficient coupling between successive reaction steps suggests that migration of the reactive oxy-ferryl ligand may be inhibited when the protein is conformationally constrained in the lattice. While several specific mechanisms outlined in the online

Supplementary Information could explain this inefficiency, all suggest a critical role for protein conformational fluctuations in modulating reaction chemistry.

Subtle changes in packing in different AlkB-ΔN11 crystal structures in space group *P*₄₃ produce substantial changes in the geometry of a putative O₂-diffusion tunnel (Fig. 2c; see also Supplementary Fig. S6). O₂ probably binds to Fe-2OG dioxygenases last after other substrates²³. The trinucleotide and 2OG seal the proximal face of the active site from solvent (as oriented in Fig. 2c), so O₂ is likely to enter via a water-filled 'O₂-binding cavity' on the opposite side that is adjacent to an ~3 Å-wide tunnel extending to the surface of the protein (at the rear of Fig. 2c and Supplementary Fig. S6B). These internal features are lined primarily by hydrophobic residues conserved in eubacterial AlKBs (Supplementary Figs S2A and S6A, B). The equivalent cavity in CAS was proposed to have a role in O₂ diffusion¹⁸ (although no opening to the protein surface was observed in that structure). The dimensions of this cavity and the adjacent tunnel vary substantially in different AlkB crystal structures due in part to conformational differences in three surface loops on the Fe-2OG dioxygenase core (Supplementary Fig. S6A). Although only a single constriction is observed in the anaerobic Michaelis complex (data not shown) and a continuous connection to the surface is observed after 2 h of air exposure (Fig. 2c), the tunnel is completely closed in the structure with Fe(II) replaced by Co(II) (Supplementary Fig. S6C). Whereas the observation of an open conformation of the tunnel supports its likely role in providing O₂ access to the active site, its variable structure suggests coupling between O₂ diffusion and protein conformational fluctuations. The restriction of these fluctuations in the lattice could inhibit O₂ diffusion into the active site and contribute to slow turnover in the crystals.

METHODS

Protein crystallization. Anaerobic crystallization was performed in an aqueous anaerobic glovebox from Mbraun, Inc. All solutions and materials except for the protein and ligand stocks were degassed and placed in the glovebox 24 h before setting up reactions. Protein, 2OG and dT-(1-me-dA)-dT solutions and Fe(NH₄)₂(SO₄)₂ crystals were brought into the glovebox immediately before use. Fe(NH₄)₂(SO₄)₂ was dissolved in degassed H₂O. The ligands were added to a 0.4 mM AlkB-ΔN11 stock solution to final concentrations of 1.8 mM for dT-(1-me-dA)-dT, 2.2 mM for Fe(NH₄)₂(SO₄)₂ and 5.4 mM for 2OG. This solution was used to set up 1:1 hanging-drop vapour diffusion crystallization reactions at room temperature. Optimized tetragonal bipyramidal crystals were obtained using a well solution containing 22–24% (w/v) PEG 3350, 10% glycerol and 200 mM sodium formate. Streak-seeding was used to control nucleation and consistently grow crystals with linear dimensions of ~100 μm. Crystals in which 2OG was replaced with succinate or Fe(II) was replaced with Mn(II) or Co(II) were grown under equivalent conditions after substituting the same concentration of the alternative ligand in the protein drop solution. Crystals grown under anaerobic conditions were oxidized by removing the tray from the glovebox and periodically lifting the cover slips to expose the drops to the atmosphere. The crystal yielding the 2 h oxidation structure was incubated with 10 mM NOC-7 (Calbiochem) during air exposure but no evidence of NO was found in electron density maps during refinement. The crystal yielding the nucleotide-free structure with succinate was grown in the presence of dT-(1-me-dG)-dT but no evidence of this ligand was found in electron density maps during refinement.

Structure determination and refinement. A three-wavelength anomalous diffraction data set was collected on crystals of the anaerobic Michaelis complex at energies of 7,118.7, 7,123.1 and 7,273.0 eV on beamline X4A at the National Synchrotron Light Source of Brookhaven National Laboratory. After data reduction and scaling in DENZO/SCALEPACK, SOLVE located the single Fe atom, and RESOLVE automatically traced 173 of 211 residues in the asymmetric unit, including side chains for 140. The structure was completed (Supplementary Table S1) using iterative rounds of manual model building in O and computational refinement and water addition in CNS. CNS was also used to refine the other structures (Supplementary Table S2) and perform occupancy refinements (Supplementary Table S3). Molecular replacement structures were solved using COMO. Ramachandran distributions for all structures are given in Supplementary Tables S1 and S2; the various structures have 88.4–93.3% of residues in core regions, 0.6% (residue Ile 201) in disallowed regions and the remainder in

allowed regions. Most molecular graphics figures were prepared with PYMOL. Literature citations for crystallographic computing and molecular graphics programs are provided in the Supplementary Information.

Received 27 October 2005; accepted 5 January 2006.

- Drablos, F. *et al.* Alkylation damage in DNA and RNA-repair mechanisms and medical significance. *DNA Repair (Amst.)* **3**, 1389–1407 (2004).
- Sedgwick, B. Repairing DNA-methylation damage. *Nature Rev. Mol. Cell Biol.* **5**, 148–157 (2004).
- Falnes, P. O., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**, 178–182 (2002).
- Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).
- Aas, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859–863 (2003).
- Ougland, R. *et al.* AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol. Cell* **16**, 107–116 (2004).
- Falnes, P. O., Bjoras, M., Aas, P. A., Sundheim, O. & Seeberg, E. Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res.* **32**, 3456–3461 (2004).
- Kataoka, H., Yamamoto, Y. & Sekiguchi, M. A new gene (*alkB*) of *Escherichia coli* that controls sensitivity to methyl methane sulfonate. *J. Bacteriol.* **153**, 1301–1307 (1983).
- Chen, B. J., Carroll, P. & Samson, L. The *Escherichia coli* AlkB protein protects human cells against alkylation-induced toxicity. *J. Bacteriol.* **176**, 6255–6261 (1994).
- Koivisto, P., Robins, P., Lindahl, T. & Sedgwick, B. Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem.* **279**, 40470–40474 (2004).
- Delaney, J. C. & Essigmann, J. M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in *alkB* *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **101**, 14051–14056 (2004).
- Falnes, P. O. Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res.* **32**, 6260–6267 (2004).
- Duncan, T. *et al.* Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl Acad. Sci. USA* **99**, 16660–16665 (2002).
- Koivisto, P., Duncan, T., Lindahl, T. & Sedgwick, B. Minimal methylated substrate and extended substrate range of *Escherichia coli* AlkB protein, a 1-methyladenine-DNA dioxygenase. *J. Biol. Chem.* **278**, 44348–44354 (2003).
- Delaney, J. C. *et al.* AlkB reverses etheno DNA lesions caused by lipid oxidation *in vitro* and *in vivo*. *Nature Struct. Mol. Biol.* **12**, 855–860 (2005).
- Elkins, J. M. *et al.* X-ray crystal structure of *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* **41**, 5185–5192 (2002).
- Muller, I. *et al.* Crystal structure of the alkylsulfatase AtsK: insights into the catalytic mechanism of the Fe(II) alpha-ketoglutarate-dependent dioxygenase superfamily. *Biochemistry* **43**, 3075–3088 (2004).
- Zhang, Z. *et al.* Crystal structure of a clavamate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centered rearrangements. *FEBS Lett.* **517**, 7–12 (2002).
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
- Mishina, Y. & He, C. Probing the structure and function of the *Escherichia coli* DNA alkylation repair AlkB protein through chemical cross-linking. *J. Am. Chem. Soc.* **125**, 8730–8731 (2003).
- Krebs, C. *et al.* Rapid freeze-quench ⁵⁷Fe Mossbauer spectroscopy: monitoring changes of an iron-containing active site during a biochemical reaction. *Inorg. Chem.* **44**, 742–757 (2005).
- Hegg, E. L. *et al.* Herbicide-degrading alpha-keto acid-dependent enzyme TfdA: metal coordination environment and mechanistic insights. *Biochemistry* **38**, 16714–16726 (1999).
- Schofield, C. J. & Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9**, 722–731 (1999).
- Wilmouth, R. C. *et al.* Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure* **10**, 93–103 (2002).
- Rohde, J. U. *et al.* Crystallographic and spectroscopic characterization of a nonheme Fe(IV)-O complex. *Science* **299**, 1037–1039 (2003).
- Que, L. Jr. The oxo/peroxo debate: a nonheme iron perspective. *J. Biol. Inorg. Chem.* **9**, 684–690 (2004).
- Price, J. C., Barr, E. W., Hoffart, L. M., Krebs, C. & Bollinger, J. M. Jr. Kinetic dissection of the catalytic mechanism of taurine:alpha-ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **44**, 8138–8147 (2005).
- Valegard, K. *et al.* The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. *Nature Struct. Mol. Biol.* **11**, 95–101 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Xiao, T. B. Acton and G. T. Montelione for provision of full-length AlkB protein, and J. C. Arbing, Y. Chen, M. Cheng, R. Farid, F. Forouhar, S. Handelman, A. Itkin, A. Kuzin, A. Petros, C. Prives, J. Schwanof, P. and W. Smith, D. Thompson, G. Verdon, W. Yong and A. Zupnick for advice and technical assistance. This work was supported by an NIH Protein Structure Initiative grant to the Northeast Structural Genomics Consortium and an Established Investigator Award from the American Heart Association to J.F.H.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession codes 2FD8, 2FDF, 2FDH, 2FDG, 2FDJ, 2FDI and 2FDK. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.F.H. (jfhunt@biology.columbia.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Valuable lessons

Most scientists would agree that graduate school packs in an incredible amount of education. But do PhD programmes teach all the right things? Most PhDs would probably say that they learned a lot about science — from research methodology to literature reviews and experimentation. But there are many other, equally important, skills that often fall through the cracks. As science and technology continue to move on, this 'soft' skill set remains more or less constant.

Soft skills include learning how to pick mentors, write grant proposals, assemble interdisciplinary research collaborations and manage labs. Without mastery of them, meeting scientific goals and climbing the ladder towards tenure can be more difficult. Most young scientists scramble to pick these skills up on their own, when they realize they need them.

Last year, I talked to hundreds of graduate students and postdocs, and I read scores of reports about the need to improve graduate education and to speed young scientists towards working independently. As a result, *Naturejobs* ran a series of articles listing soft skills and showing how to

acquire them. This week, that entire series, together with additional relevant articles, has been posted in the Editor's Choice part of the *Naturejobs* website (www.naturejobs.com).

I hope that this will encourage administrators to incorporate some of these skills into graduate education. But if they don't, then at least young researchers should be able to pick up what they need to know faster, on their own.

In an ideal world, administrators would not only ensure that graduate students get this sort of training, but that undergraduates are aware of the skills. Because undergraduates who join graduate school knowing how to pick mentors and speed through the system will have a more productive time focusing on science — making the most of what graduate school already does well.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building,
 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

European Sales Manager:
 Andy Douglas (4975)
Natureevents: Sille Opstrup (4994)

UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
**Germany/Austria/
 The Netherlands:**
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

**European Satellite Office
 Germany/Austria/
 The Netherlands:**
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Drug hunt

In the first of a series of articles on the drug pipeline, **Hannah Hoag** looks at the opportunities in selecting leads.

If industries were gardens, the biopharmaceutical sector would be uneven and rocky, with only the most persistent plants thriving in an often unpredictable environment. Last year was particularly rough: market withdrawals of blockbuster drugs, the looming end of patent protection on key products, and constant reminders of nearly empty pipelines. Some companies have announced significant layoffs over the next few years; others look set to outsource some services to lower-cost locations. But in terms of employment, which stages of the drug pipeline stay steady? For the time being, chemical and biological scientists eager to get their hands wet in drug discovery can rest easy.

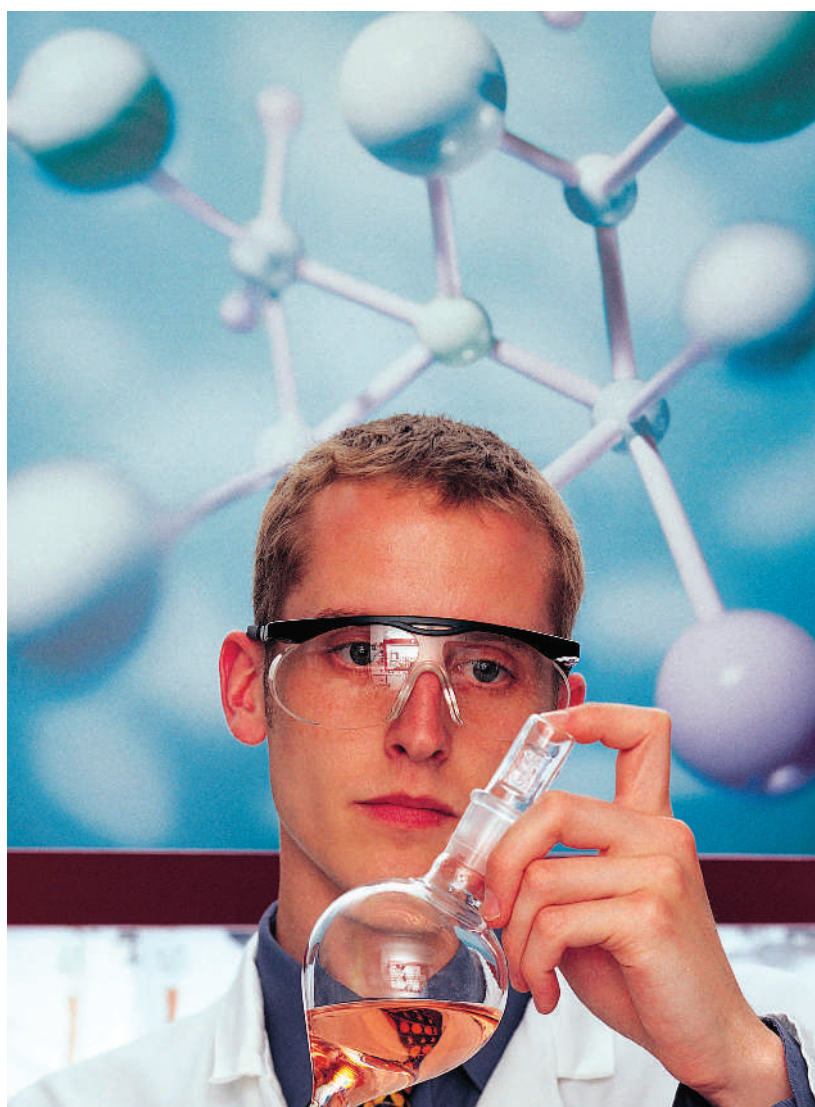
Recruitment has been slow at biotechnology and pharmaceutical firms for the past couple of years. But companies continue to be on the lookout for talent that can bring new drugs into the pipeline.

"We always recruit year by year, even through the thin years," says Martin Braddock, associate director of discovery bioscience at AstraZeneca, UK. According to the US Bureau of Labor Statistics, employment for chemists and biologists in the US pharmaceutical and medicine manufacturing sector is expected to grow by about 25% until 2014. Jobs still exist at the early stages of the drug pipeline, but many have been redistributed and may be more competitive than ever.

Hitting the target

In the earliest stages of pharmaceutical research, scientists are part of medical reconnaissance missions to identify biological targets that could be attacked by a new drug. Biologists play a much larger role than chemists during this stage. They track down targets using literature searches, database mining and state-of-the-art techniques, including antisense and small RNAs to reduce the expression of targeted genes.

The biologists who bring these targets to light are often plucked directly from academia. "We don't necessarily recruit for expert skills. What we really



ASTRAZENECA

Researchers with a history of problem-solving are highly prized for early-stage drug discovery.

look for are broad-based scientists who have shown a history of problem-solving in a biological system," says Gary O'Neill, senior director of biology at Merck Frosst, in Montreal, Canada.

With the abundance of data now available from genome mapping, work at this stage is heavily supported by those with experience in bioinformatics to sift through the information. Structural chemists may also get involved in providing information about the three-dimensional shape of the target, says Gerd Schnorrenberg, head of chemical research at Boehringer Ingelheim in Biberach, Germany. If X-ray crystallography or NMR spectroscopy aren't possible, computational chemists are called on to produce a three-dimensional model of the target.

During the lead identification stage, drug candidates are found in a number of ways. These days, drug leads rarely come from natural products. Instead they're found within proprietary databases containing hundreds of thousands, or even millions, of compounds, and the scientific and patent literature. These high-throughput screens give a discovery team a hit list of compounds that show interactions with the target. The compounds are then clustered using bioinformatics tools. "These scientists need a broad knowledge of molecular biology, cell biology and automation," says Schnorrenberg.

THINK SMALL

Between 1990 and 2004, the American Chemical Society tracked chemistry graduates to see whether they wound up in small or large industry. Over the years, the changing results have proved significant. In 1990, more than 70% of graduates were working for large employers, but by 2004 the distribution was nearly 50:50.

Working for a large pharmaceutical firm has its advantages, as the vast in-house experience provides plenty of learning opportunities. But you may see only a small part of the programme. Smaller companies provide the opportunity to focus on a wide variety of skills, techniques and protocols.

"If you are young, bright, hardworking and ambitious, you can progress rapidly in biotech," says David Campbell (left), now vice-president of drug discovery at Phenomix, who previously worked at Bayer. "If you can step into that vacuum and add value and drive programmes, you can be promoted very rapidly."

H.H.



R. STARKMAN

ASTRAZENECA

The essential skills required for lead generation are found in medicinal chemistry. "Making molecules is what it's all about," says Paul Leeson, director of medicinal chemistry at AstraZeneca. The task at this stage is to design molecules with appropriate activity against the target, and selectivity against the counter-targets, to improve the potency of the hits generated during screening. New graduates or postdocs keen to pursue organic synthesis can find an entry point into the lead identification stage, says Leeson. The medicinal chemistry can be taught on the job or through short courses (see *Nature* 424, 594–596; 2003).

Structural chemists and those with computational skills will also find jobs in lead identification. Their ability to determine molecular structures using X-ray crystallography or NMR, or to pick out salient information from structural databases, can help medicinal chemists design new drugs. Although there are fewer positions available, chemists with these skills may be recruited directly out of graduate school.

Smaller biotech firms, which frequently don't have access to vast chemical libraries for high-throughput screening, can jump-start their programmes by focusing on families or subfamilies that are likely to produce good leads, says David Campbell, vice-president of drug discovery at Phenomix in San Diego, California. "You need a good medicinal-chemistry background, and knowledge of what sorts of molecules have worked in the past against the targets and which ones have a higher potential of having drug-like properties," he says.

Biologists who hope to break into drug discovery by generating leads for new drugs need to like working with cell-based systems and to be able to improve an already well-oiled machine.

Lead optimization

Moving a lead to the optimization stage requires greater investment. This step refines compounds to see whether they are worth developing further. At this point, the number of leads has been boiled down to a few: a primary lead plus a couple of back-ups. Pharmacologists and scientists with training in pharmacokinetics and metabolism work with medicinal chemists during this stage, testing and tweaking compounds until the optimal molecule is made.

Companies are putting increased emphasis on a lead compound's absorption, distribution, metabolism, excretion and toxicology profile during the lead optimization stage. These data help project leaders identify — and if necessary kill — lacklustre drug candidates earlier in the pipeline.

L. LAUTH

At deCODE Chemistry in Woodbridge, Illinois, the drug-metabolism team has doubled in size in the past year. "We see that as being a real bottleneck," says David Zembower, vice-president of chemistry. "It's an area where you separate the interesting chemicals from the potential drugs."

Chemists would be well-advised to look for jobs at small and medium-sized service-oriented companies (see "Think small"). Limited capital is forcing small biotech firms to outsource aspects of their discovery operations, meaning that service jobs are available. "The wisest use of their cash is not building lab facilities and hiring more people, it's outsourcing their science to companies such as us to create intellectual property and further value," says Zembower.



Pushing drug candidates through the discovery pipeline requires a range of skills.



AMRI

And when it comes to intellectual property, patent lawyers tend to be involved at the earliest stages of drug discovery. "It has become more important than it used to be," says Schnorrenberg. "In the past it was mainly dedicated to chemistry patents; now it is also going into targets. The strategy is more aggressive and it's a higher workload."

Patenting a novel target occurs early on, but patenting compounds is done as late as possible so that the patent will last longer. It's risky, as it could leave the way open for a competitor. Lawyers involved in these non-research aspects of drug discovery will almost always have experience in biology or chemistry.

The relationship between start-ups and larger firms is also changing. Rather than scooping up the best leads through mergers and acquisitions, large firms are contracting out an increasing proportion of drug discovery to smaller outfits.

Both deCODE Chemistry and Albany Molecular Research (AMRI) of Albany, New York, say they are planning for some growth in 2006. AMRI plans to hire PhD-level medicinal chemists and bachelor's or master's-level physical chemists at its US facilities this year, says Bruce Sargent, vice-president of discovery research and development. AMRI also recently set up subsidiaries in India and Singapore. Although still in their infancy, these are growing fast to provide resources in discovery and synthesis to the US branch. As such strategic outsourcing gains steam — fuelling job growth in India, China and the former Soviet Union (see *Nature* 433, 902–903; 2005) — North American and European firms may have fewer jobs to offer.

Interaction among scientists is ubiquitous in drug discovery, meaning that strong communication and interpersonal skills are essential. Increasingly, scientists need to have an understanding of a discipline outside their own. Project leaders need to keep the final goal in mind: what does the customer, the clinician, the principal investigator or the patient want in a drug?

Flexibility is another important trait, especially in the biotech industry, says Campbell, where a company can change direction and focus several times. Project leaders need to be able to carry a team through the highs and lows. "Keeping momentum is key for a project leader," says Braddock.

Hannah Hoag is a freelance writer based in Montreal, Canada.



David Zembower sees assessing drug metabolism as a pipeline bottleneck.

MOVERS

Giulio Superti-Furga, scientific director and chief executive, Research Center for Molecular Medicine, Austrian Academy of Sciences



2000–04: Co-founder, scientific director (2000), vice-president of biology (2001), senior vice-president of biology (2003), Cellzome, Heidelberg, Germany
1995–2004: Team leader, European Molecular Biology Laboratory, Heidelberg, Germany

Five years ago, Giulio Superti-Furga jumped from the tranquillity of academia into the high-risk, rough-and-tumble world of biotechnology. After an exciting but exhausting ride, he has recently returned to academia and is now combining the best of both worlds in his new lab in Vienna.

A molecular biologist, Superti-Furga got his first taste of industrial research as a PhD student, spending a year at Genentech in California. He went on to do a postdoc at the European Molecular Biology Laboratory (EMBL) in Germany, where he stayed for the next 14 years studying the proteins involved in cell signalling.

Near the end of the 1990s, Superti-Furga began thinking about launching his own company, but wasn't sure where to begin. Then he met some biotech entrepreneurs, who had come to EMBL scouting for ideas and people for their next start-up. This turned out to be one of the most important meetings of his life. "It was like playing guitar in your backyard and then the stretched limo arrives and The Rolling Stones step out and ask you to play with them," he says.

He co-founded Cellzome, a proteomics company that would both collaborate with drug firms to find better biological targets and develop its own drugs. Superti-Furga worked one day a week at EMBL while serving as scientific director and later as vice-president at Cellzome.

Suddenly, Superti-Furga found himself tackling a variety of new tasks, such as pitching his technology to dozens of investment bankers and venture capitalists. He found this audience far tougher than the most sceptical scientists. "In business, if they think you're a fool, they tell you straightaway," he says.

By 2004, Cellzome was better established and Superti-Furga was ready to move on. So when he got the chance to head up the new Center for Molecular Medicine at the Austrian Academy of Sciences, he didn't think twice. Admittedly, he had promised his Viennese wife that he would apply for jobs in her home town, but he also saw an opportunity to build another new organization.

Now he is using lessons learned at Cellzome: half of his lab consists of scientists from industry. He is hoping to combine the free-wheeling idea generation of academia with the efficient and professional execution of ideas characteristic of the biotech industry. Throwing together people from different backgrounds is consistent with his advice to young scientists: "Try to associate with people who are smarter than you."

Corie Lok

SCIENTISTS & SOCIETY

The changing face of Japan

Many young scientists in Japan find the academic environment frustrating and rigid because veteran researchers are well entrenched and career prospects seem limited.

To address this problem, the national government this year plans to introduce a few programmes to support young scientists. "So far, our investment has been focused on facilities and equipment," says Masaaki Tanaka, director of the knowledge infrastructure policy division at the education ministry. "Now, it's time to shift our focus to invest in people."

One key programme aims to encourage universities and research institutes to introduce a tenure system. A standard fixture at US universities, tenure has yet to be adopted in Japan. Instead, the country largely runs a system based on seniority in which young researchers end up working for their supervisors once they have completed their postdoc or stay in uncertain contract positions for many years. A tenure system promises a higher-level, more independent, permanent position if a researcher works for a certain period as a contract worker and generates good results during that time.

To fund this programme, the education ministry plans to give between ¥200 million (US\$1.7 million) and ¥300

million every year for up to five years to each of about 30 universities and research institutes. The money would be distributed to young researchers in their 20s and 30s chosen to fill tenure-track positions. These researchers would be independent of professors or assistant professors and have their own labs, equipment, budgets and assistants.

The new system "will certainly provide incentives for young researchers", says Hiroaki Suga, a chemical biologist at the University of Tokyo and an advocate of improving Japan's research and education. But he points out that it won't work well unless a system to evaluate the researchers' work properly is established.

Another programme being introduced by the government is targeted at young female researchers. One of its aims is to encourage women to return to work after maternity leave. Currently, many female postdocs and assistants can't get any benefits for childcare so they quit their jobs. With this programme, about 30 selected women researchers will receive about US\$3,100 every month for two years to help pay for childcare and help them continue their research. "We want talented female students to join the science world," Tanaka says.

Ichiko Fuyuno is a contributing correspondent for Nature in Tokyo.

GRADUATE JOURNAL

The end of the affair

If you want a PhD topic that will amuse your friends endlessly, try mine. For the past three years, I've worked with ants as a model of social behaviour — feeding ants, doing ant genetics and counting ants by the thousands. Graduation is a year away and writing up feels like a mountain made out of an anthill. But it pales in comparison to the biggest challenge I face this year: choosing a new career.

Over the course of my graduate studies, I've slowly realized that basic research isn't for me. As much as I agree with the idea of science for science's sake, my dream day at the office would produce more tangible results.

I've recently enrolled in a year-long mentoring programme where I meet people who have made the leap from academia to the real world. It has helped me understand that the skills gained in a PhD can be applied outside the lab as well. At the moment, my ideas range from journalism to pest management in Africa.

For someone who has focused on basic science since her undergraduate days, making the decision to quit has been like the slow death of a relationship. First, you pretend everything is all right; then, you wonder what is wrong with you; and, finally, you are forced to acknowledge that what you have invested so much time and energy in may not be what you really want. The period of mourning is over. I'm ready to move on, towards a whole new world.

Katja Bargum is a graduate student in evolutionary biology at the University of Helsinki in Finland.

Daddy's slight miscalculation

Science at the cutting edge.

Ashley Pellegrino

Zzzzzzt! Clink! "Ouch!"

I listened from outside Dad's lab. He was working on a new experiment.

No one knew Daddy had started a new project, except me. Only I know when he is working on an experiment, because he usually works while my brother and sister and I are asleep. Whenever he is up all night, I know he's working on something new, sometimes even something strange — like the pink octopus whose ears were like wings.

He doesn't go out to work in the morning like other fathers. Did he ever go out to work, like other kids' dads? I think so; but he does not like to talk about the time before I was born. Something happened. Something sad, I think.

Except for us, Dad has only his work.

If you watch science fiction, or read it, you might think it's boring or even scary growing up with a scientist in the house. Actually, it's more interesting than scary, more easy than boring. When I ask for help with my homework, he sometimes gets me all dizzy. Other times, it becomes crystal clear. But don't ask him why nature is the way it is, unless you really want to know, unless you want to get dizzy trying to understand how black paint looks black because it is really trapping all the other colours inside.

"Why is the sky blue?" I remember asking.

Most adults would answer: "Go ask your science teacher." Or: "That's how God made it."

But Dad said: "That's the colour the sky rejects."

I don't completely understand; but I'm sure not many kids walk around wondering about living in rejected colours. It makes me wonder a lot about the things around me. It's actually kind of fun. Even with ideas gone completely strange, we're always entertained. None of us will ever forget what Dad did to Mr Kitty.

Cat's toes and monkey fingers?

Mr Kitty was already nicknamed Mr Mischief, even without Dad's help.

'Biomorphing' paws into hands — for a cat? Who wouldn't have guessed that

could lead to chaos? Just imagine living with a cat that can unscrew jars and open doors. But even if our house did eventually end up with a child lock on the refrigerator door (although there were no toddlers around), it was really fun, believe it or not, to have a cat with hands.

Yet the noises that came from Dad's lab, that night — they sounded more creepy than fun. Usually the sounds coming from Dad's lab are buzzes, not ouches.

So, I knocked.

"Come in," he said.

The sun was shining through his windows and

"But Dad. What happens when...?"

"Don't worry," he said. "Nothing can go wrong. Nothing goes wrong in science, if you really do your homework. Science only improves life."

"Isn't that what they said about the *Titanic* and the *Columbia*?" I asked.

"Ouch!" Dad said. "Gotta think about that one. But this science is different. Trust me."

Dad was changing. A lot of people were changing, these days. Some were changing for the better. Some, for the worse.

Dad always used to tell me that the most important place for a scientist to be is in the unknown. "I don't know," he had taught me, "is always a good place to start."

Now, he was starting to speak as if he did know — as if, these days, he was afraid to say the words "I don't know".

I thought again about those angry trees in *The Wizard of Oz*, and I did not like the whole idea. "Dad, wait..."

But Dad wore the wire anyway, and he flipped the switch, saying: "This is something new. This is something completely different."

Just then, my brother Kyle went into the yard, wanting only to do something nice for Dad. He wanted to give Dad a surprise, and to do some of the household chores for him. Kyle was heading for the grass. He was pushing Dad's little mechanical lawnmower. He was pushing fast, really fast.

Dad never knew that a single blade of grass could scream. No one knew, until that day. Kyle mowed 20 feet, before Dad's screams stopped him.

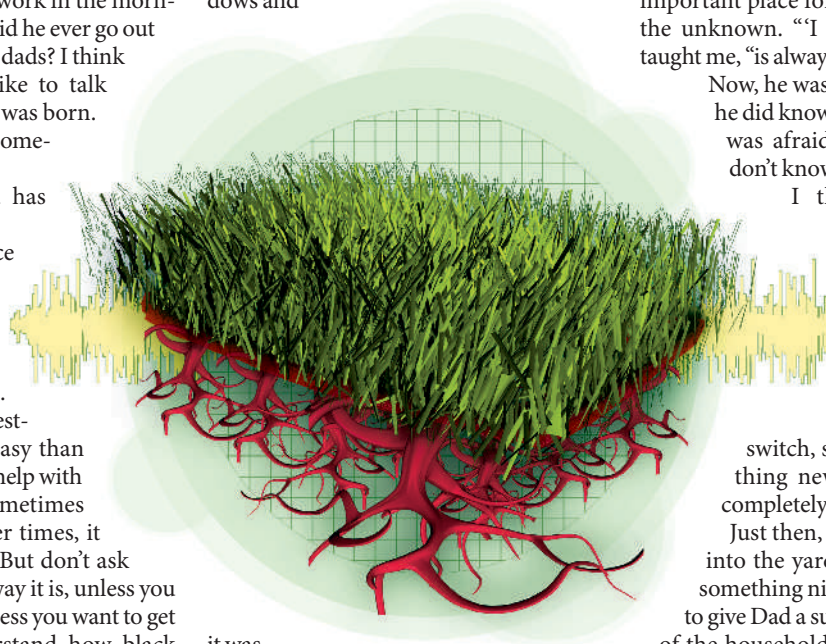
Yes, this was as different as Dad promised it would be.

The grass on our lawn is taller than me, now. And the neighbours have begun to complain. But Dad will not cut the grass; not a single blade.

Dad still won't tell me what it sounded like.

And I wonder.

Ashley Pellegrino is the 11-year-old daughter of a scientist. Aside from writing tales that occasionally scare the scientist whose real-life adventures have been known to scare Stephen King, she's a perfectly typical sixth-grader.



it was a bright new day outside. Dad had worked all night. This had to be something special. I saw what might have been a small device used to communicate with those pink octopuses with Dumbo's ears.

"Dad, what is that?" I asked.

"Bone-phone," Dad replied. There was a strange-looking wire clipped to his ear.

"I don't like the sound of that," I said.

"That's OK. We can always name it again."

"Fine! But, what does it do?"

"Do you see those orchids on the table? And those trees outside? This wire sends a signal right into my brain. It lets me feel exactly what the plants feel. Isn't that amazing?"

I looked at the fruit trees, and I looked at the tall green grass. I thought of *The Wizard of Oz*. I thought of how the trees yelled at Dorothy when she picked their apples.

JACEY